



UNIVERSIDADE NOVA DE LISBOA  
INSTITUTO DE HIGIENE E MEDICINA  
TROPICAL



MESTRADO EM CIÊNCIAS BIOMÉDICAS

**POLIMORFISMOS DA REGIÃO DA INTEGRASE,  
DO GENE *POL*, EM DOENTES INFECTADOS POR  
VIH-1**

**JOAQUIM LUÍS SOARES CABANAS**

**2010**



UNIVERSIDADE NOVA DE LISBOA  
INSTITUTO DE HIGIENE E MEDICINA  
TROPICAL



# **POLIMORFISMOS DA REGIÃO DA INTEGRASE, DO GENE *POL*, EM DOENTES INFECTADOS POR VIH-1**

**JOAQUIM LUÍS SOARES CABANAS**

*Tese apresentada para a obtenção do grau de  
Mestre em Ciências Biomédicas.*

**Orientador:**  
**Prof.<sup>a</sup> Doutora Perpétua Gomes**  
**Co-Orientador:**  
**Prof. Doutor Celso Cunha**

**2010**

## Agradecimentos

Quero agradecer a todas as pessoas que me ajudaram e tornaram possível a realização deste trabalho:

- À Professora Doutora Perpétua Gomes a orientação, a disponibilidade, a paciência e as sugestões.

- Ao Professor Doutor Celso Cunha, as opiniões e as sugestões.

- Ao Dr. Ricardo Camacho pela ajuda na escolha do tema.

- À Doutora Ana Abecasis, pela ajuda importantíssima no tratamento estatístico dos dados.

- Aos meus colegas de laboratório Fátima Gonçalves, Inês Costa. Um especial agradecimento à Patrícia Carvalho e à Isabel Diogo pela colaboração no estudo realizado.

- Ao meu cunhado Eugénio pela revisão ortográfica.

A todos os profissionais do serviço de medicina transfusional do CHLO-HEM o meu obrigado pelo saudável convívio.

Finalmente agradeço aos meus pais, à Fátima, ao Dinis e ao Diogo por tudo e mais alguma coisa.

A todos muito obrigado.

## Índice

Agradecimentos	I	
Índice de figuras	IV	
Índice de tabelas	V	
Glossário	VI	
1	Capitulo I-Introdução	3
1.1	Breve evolução terapêutica	3
1.2	Factores de resistência aos anti-retrovirais em doentes VIH-1	5
1.3	A heterogeneidade genética do VIH-1: um impedimento ao sucesso terapêutico dos anti-retrovirais	7
1.4	A influência da recombinação genética na diversidade do VIH-1	7
1.5	. Pressão selectiva imposta pelos anti-retrovirais.	8
1.6	Estrutura da integrase	8
1.7	Região N-terminal	10
1.8	Região Central	13
1.8.1	Características da Região Central:	14
1.9	Fixação do Mg <sup>2+</sup> ao centro catalítico	18
1.10	Região C-Terminal	19
1.11	Actividade catalítica da integrase	21
1.12	A importância do complexo de pré-integração	23
1.13	A importância do Mg <sup>2+</sup> nas reacções catalíticas da integrase	25
2	Capítulo II - Inibidores da integrase	26
2.1	Introdução.	26
2.2	Mecanismo de acção dos inibidores	26
3	Capitulo III - Polimorfismos da integrase	31
3.1	Introdução	31
3.2	Objectivos	32
3.2	Material e métodos	32

3.2.1	Amostras	32
3.2.2	População alvo	32
3.2.3	Extracção do ARN viral	33
3.2.4	Protocolo de amplificação	33
3.2.5	Quantificação dos produtos da RT-PCR	35
3.2.6	Purificação dos produtos amplificados	36
3.2.7	Reacção de sequenciação dos produtos purificados do gene da integrase.	37
3.2.8	Detecção e leitura da sequência de ADN.	39
3.2.9	Edição das sequências obtidas.	40
4	Resultados	41
4.1	Análise da região N-terminal.	43
4.2	Análise da região central.	43
4.3	Análise da região C-terminal	45
4.4	Análise geral	46
<b>5</b>	<b>Co-variação do subtipo G do VIH-1</b>	46
5.1	Importância do subtipo G em Portugal	46
6	Discussão	51
7	Bibliografia	55

## Índice de figuras

Figura 1:Ciclo replicativo do VIH-1 e alvos terapêuticos dos anti-retrovirais.	6
Figura 2: O gene <i>pol</i> do provirus do VIH-1	9
Figura 3: As regiões da integrase	10
Figura 4: Diferentes, estruturas dedos de zinco.	11
Figura 5:Região N-terminal isolada e ligada a região central	12
Figura 6:Ligação do Zinco ao arranjo HHCC	12
Figura 7:Estrutura dimérica da região Central	14
Figura 8:Organização da região Central	15
Figura 9:Laço flexível do Centro catalítico	17
Figura 10:Estrutura da região central na proximidade da tríade catalítica da integrase	18
Figura 11:Ligação do $Mg^{2+}$ à tríade catalítica	19
Figura 12:Ligação da região Central à região C-terminal	20
Figura 13:Ligação da região C-terminal através do domínio SH3 ao ADN	20
Figura 14:Sequência de eventos da integrase e da transposase	23
Figura 15:Complexo de pré-integração	24
Figura 16:Estrutura molecular de (DKA) ácidos diceto-aril	28
Figura 17:Mecanismo de ligação dos inibidores à integrase	29
Figura 18:Aminoácidos envolvidos na resistência aos DKA	30
Figura 19:Exemplo de preenchimento de uma placa para sequenciação	38
Figura 20:Edição de uma sequência	40
Figura 21:Alinhamento da sequência de referência HXB2	41
Figura 22:Fotografia das bandas dos produtos amplificados	42
Figura 23:Resultados percentuais da variabilidade das posições dos aminoácidos	44

## Índice de tabelas

Tabela 1: <i>Master mix</i> para uma amostra	33
Tabela 2:Ciclos de amplificação da RT-PCR da integrase	34
Tabela 3:Tabela de referência para a quantificação dos produtos de RT-PCR	35
Tabela 4:Tempos utilizados na purificação	36
Tabela 5:Valores de diluição para produtos de RT-PCR purificados	36
Tabela 6: <i>Primers</i>	37
Tabela 7:Ciclos utilizados na amplificação do gene da integrase	38
Tabela 8 :Valores de associação Jaccard	49

## Glossário

ADN- ácido desoxirribonucleico.

ARN- ácido ribonucleico.

ADN circular- ADN viral não integrado

CPI- complexo pré-integração do ADN viral.

*Docking* – local de ancoragem do inibidor.

Dka- ácido diceto áril.

EDTA- ácido etil-diemina-tetra-acético, anticoagulante.

Emerina- transportador localizado nos poros nucleares.

Electroferogramas- picos gráficos coloridos representando as quatro bases do ADN.

Fasta - ficheiro informático que contem a sequencia em estudo.

*Fitness*- aptidão, expressão usada em estudos de evolução.

Grupo M- grupo filogenético Major do VIH-1.

GeneBank- *website* onde são encontradas as sequências de referência validas.

H AART- regime terapêutico de alta potência.

IP- fármaco inibidor da protease.

Integrase- enzima que catalisa a integração do ADN viral.

Jaccard- teste estatístico não paramétrico.

Jack-knifed- método estatístico para validar o índice de Jaccard.

LEDGF/p75- factor de crescimento *lens epithelium*.

*Loading buffer*- bromofenol.

Master mix- mistura química usada na utilização de uma PCR.

NRTI- fármaco inibidor da transcriptase reversa.

NNRTI- fármaco inibidor da transcriptase reversa não análogos dos nucleósidos

*Primer* – sequencia oligonucleótídica iniciadoras.



Pop 6- polímero condutor.

Polimorfismo- variantes aos aminoácidos habitualmente encontrados.

Raltegravir- fármaco inibidor da integrase

RMN- ressonância magnética nuclear.

Spss – *Statistical Package for the Social Sciences*.

*Turnover*- taxa de substituição.

TBE- Tris/Borato/EDTA

*Wild-type*- estirpe original.

## Resumo

A integrase do VIH-1 é a enzima responsável pela integração do ADN viral no ADN da célula hospedeira. Este processo é indispensável à replicação viral, desenvolvendo-se em duas etapas independentes, o processamento 3' e a transferência de cadeias catalisadas pela integrase. A compreensão das interacções entre a integrase e o ADN e a cinética da formação dos complexos de pré-integração evidenciaram a necessidade de descobrir inibidores das ligações do complexo integrase/ADN. Esta necessidade, permitiu desenvolver o Raltegravir e o Elvitegravir como inibidores potentes da transferência de cadeias e consequentemente da replicação viral *in vivo*. Estes fármacos porém, não escapam aos fenómenos de resistência. Mais de 40 substituições na integrase foram associadas ao desenvolvimento de resistência ao Raltegravir e/ou elvitegravir, *in vitro*. Algumas destas alterações foram também encontradas *in vivo* em doentes com falência terapêutica aos inibidores da integrase. Os polimorfismos do gene da integrase ocorrem de forma natural, podendo ter implicações importantes no desenvolvimento de inibidores da integrase. Foi objectivo deste estudo a determinação da prevalência de polimorfismos em doentes não tratados com inibidores da integrase e a frequência de substituição de aminoácidos que ocorrem de forma natural em posições associadas a resistência. Foram indentificadas as posições polimórficas do gene *pol* da integrase em doentes não tratados com inibidores da integrase, confirmando-se a ausência de mutações associadas a resistência nestes mesmos doentes. A percentagem de polimorfismos variou de acordo com a região analisada, sendo a região C-terminal a mais variável. A análise mais detalhada do subtipo G contribuiu de forma significativa para caracterizar os polimorfismos existentes na população de doentes VIH-1 portuguesa. A utilização do índice de Jaccard permitiu de forma inequívoca demonstrar a associação entre posições polimórficas e verificar como essas associações poderão evoluir.

## Abstract

The integrase catalyzes the integration of HIV-1 DNA into the host cell chromosome in a process which is essential for viral replication. Integrase performs two independent reactions, 3'-end processing and strand transfer. Deciphering the structural determinants of the interaction between integrase and its substrates as well as kinetics of this interaction shed light on the importance for pharmacological agents to target the integrase-viral DNA complex. This approach led to the development of Raltegravir and elvitegravir which turned out to be powerful inhibitors of strand transfer both *in vitro* and *in vivo*. However, like others ARVs, these compounds face resistance phenomenon. Among them, about 40 substitutions *in vitro* have been specifically associated with the development of resistance to Raltegravir and/or Elvitegravir, some of them were also found *in vivo* in patients failing such integrase inhibitors. Naturally occurring integrase gene polymorphisms may have important implications for integrase inhibitors development. This study evaluated clinical HIV-1 strains from integrase inhibitors naïve patients, to determine the prevalence of integrase gene polymorphisms, and the frequency of naturally occurring amino acid substitution at positions associated with integrase resistance. Polymorphic positions at integrase region of *pol* gene were identified in patients not treated with integrase inhibitors. The absence of resistance mutations to these compounds was confirmed. The percentage of polymorphisms was different within integrase regions, being the C-terminal region the one with the highest variability. Detailed analysis of G subtype contributed significantly to characterise the polymorphisms in Portuguese HIV-1 infected patients. Jacard index was used to demonstrate association between polymorphic positions and verify how those associations may evolve.

# 1 Capítulo I-Introdução

## 1.1 Breve evolução terapêutica

O primeiro relato da infecção pelo vírus da imunodeficiência humana tipo 1 (VIH-1), nos Estados Unidos, com a identificação de pneumonias a *Pneumocystis carinii* e sarcoma de Kaposi, entre homossexuais aparentemente saudáveis, aconteceu há 30 anos. (Gottlieb *et al.* 1981). O uso de anti-retrovirais e a sua utilização na prática clínica acompanharam a evolução da infecção pelo VIH-1 ao longo deste longo período. Longe vão os tempos da utilização de Zidovudina (AZT) em regime mono terapêutico. O uso combinado de vários anti-retrovirais em regime de alta potência *highly active antiretroviral therapy HAART* e a descoberta de novos alvos terapêuticos, marcam a actualidade (Figura 1).

As mutações sofridas pelo VIH-1 são o maior entrave ao sucesso terapêutico com anti-retrovirais e ao desenvolvimento de vacinas. Os regimes terapêuticos *HAART* permitem que a supressão viral por longos períodos seja uma realidade. No entanto, existem ainda alguns desafios por vencer no combate à infecção, nomeadamente, resistência, toxicidade, capacidade de penetração dos anti-retrovirais, adesão aos regimes terapêuticos, eliminação da replicação viral nos reservatórios celulares e aumento da capacidade de resposta imunológica do hospedeiro.

Os regimes *HAART* são regimes terapêuticos de longa duração, que levam à selecção de estirpes VIH-1 com mutações. A acumulação de mutações aumenta o número de resistências aos anti-retrovirais, actualmente utilizados na terapêutica:

- Inibidores da transcriptase reversa análogos dos nucleósidos/nucleótidos (NRTI),

- Inibidores da transcriptase reversa não análogos dos nucleósidos (NNRTI),
- Inibidores da protease (IP).

Este aumento de resistências leva conseqüentemente a um aumento da carga viral, com eventual progressão da doença.

Na prática clínica, as opções farmacológicas para os doentes em falência terapêutica, têm de ser cada vez em maior número.

O VIH-1 pertence ao grupo dos retrovírus, contém tipicamente 3 à 6 genes acessórios/regulação adicionais (Potter et al. 2004). No caso do VIH-1, os acessórios são: *vif*, *vpr*, *vpu*, e os de regulação: *nef*, *tat*, *rev*. Apesar da complexidade genómica, apenas um número limitado de proteínas essenciais ao ciclo de replicação do VIH-1, são actualmente alvos terapêuticos (Figura 1). O VIH-1 apresenta uma grande diversidade genética, com origem em factores, tais como: erros cometidos pela transcriptase revers, rápido *turnover* viral *in vivo*, pressão selectiva do hospedeiro, recombinação genética. Devido a esta diversidade, as variantes do VIH-1 são classificadas em três grupos filogenéticos principais: grupo M (major), O (*outlier*), N (nãoM/nãoO) e mais recentemente foi identificado mais um grupo filogenético- o grupo P (Vallari et al. 2010). O grupo M é constituído por 10 subtipos filogenéticos (A-K). As combinações genéticas entre subtipos diferentes originam as formas circulantes recombinantes CRFs (Buonaguro et al. 2007).

A terapêutica *HAART*, inclui no mínimo a combinação dois NRTI, com um inibidor da protease IP e/ou um NNRTI. O aprofundamento de conhecimentos sobre a estrutura biológica do VIH-1 e as respectivas interações com a célula hospedeira é fundamental para

permitir a descoberta de novos alvos terapêuticos. Entre os diferentes alvos virais já estudados para o desenvolvimento de novos fármacos, os mais recentes incluem:

- Inibidores de entrada,
- Integração do ADN,
- Inibidores de maturação de partículas virais.

Actualmente, existem aproximadamente 25 fármacos anti-retrovirais, utilizados na terapêutica da infecção pelo VIH-1, e várias recomendações elaboradas por peritos, que visam a utilização combinada destes anti-retrovirais. (Hammer *et al.* 2008).

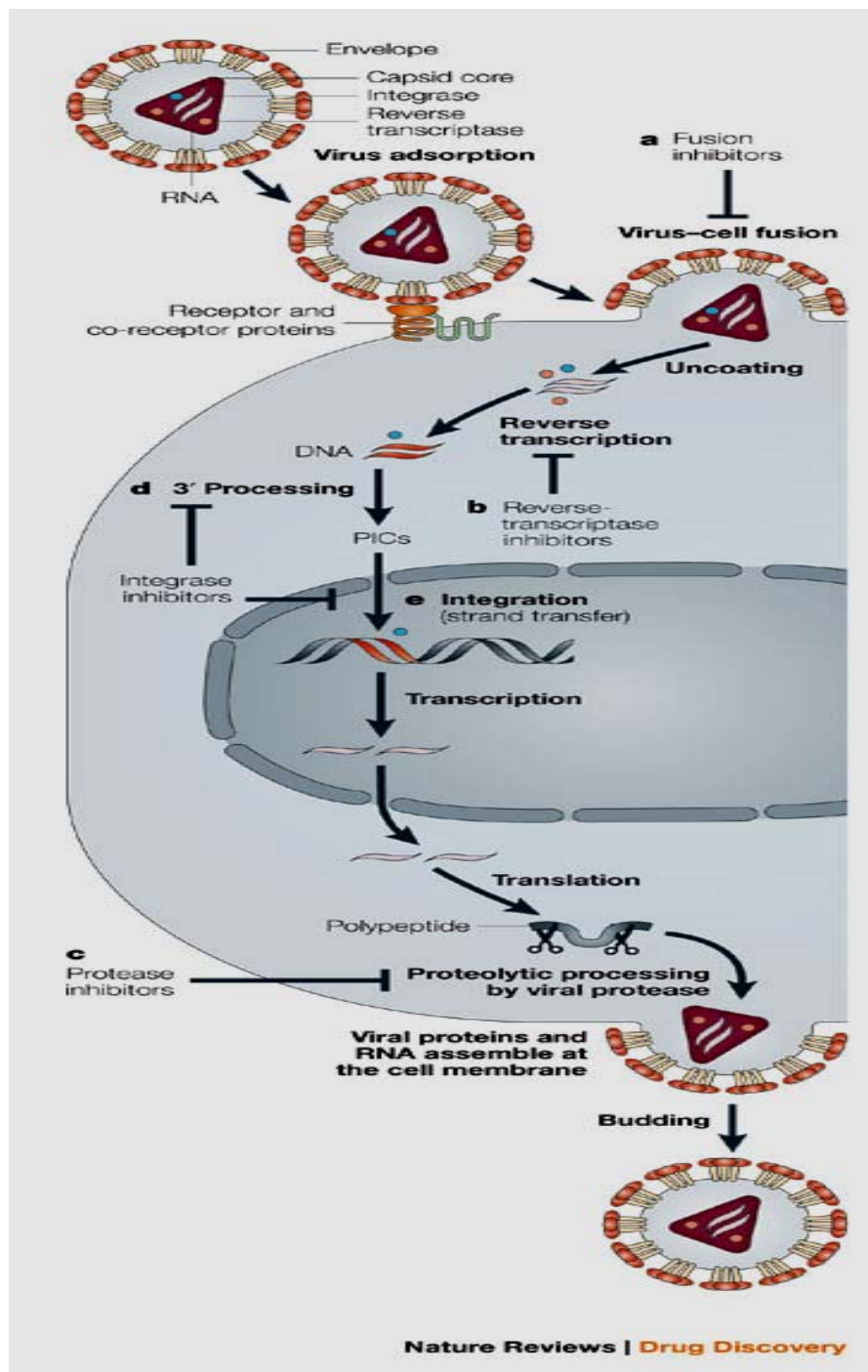
A utilização destas novas classes de anti-retrovirais, em associação com as já existentes, proporcionará uma inibição muito mais potente no aparecimento de resistências aos anti-retrovirais.

## **1.2 Factores de resistência aos anti-retrovirais em doentes VIH-1**

Apesar do decréscimo da mortalidade e morbilidade e de um aumento significativo da esperança média de vida, associada aos regimes *HAART*, em indivíduos infectados pelo VIH-1, a eventual falência terapêutica e progressão para SIDA mantêm-se praticamente inevitáveis. (Potter *et al.* 2004).

O aparecimento de resistências aos anti-retrovirais, em indivíduos sob terapêutica *HAART*, pode ter origem em diferentes factores:

- Elevada taxa de mutações,
- Elevada taxa de replicação viral,
- Pressão selectiva proveniente da terapêutica,
- Farmacocinética,
- Má adesão aos regimes terapêuticos,
- Existência de reservatórios virais (SNC, aparelho reprodutor masculino).



**Figura 1:**Ciclo replicativo do VIH-1 e alvos terapêuticos dos anti-retrovirais.

*Adaptado de:* [www.nature.com/.../v4/n3/fig\\_tab/nrd1660\\_F1.html](http://www.nature.com/.../v4/n3/fig_tab/nrd1660_F1.html)

### **1.3 A heterogeneidade genética do VIH-1: um impedimento ao sucesso terapêutico dos anti-retrovirais**

A variabilidade molecular do VIH-1 assenta em erros cometidos pela transcriptase reversa (TR). Os vírus ARN, incluindo os retrovírus como o VIH-1, têm taxas de mutação aproximadamente 100 vezes superiores às dos vírus ADN, bactérias ou organismos eucariotas (Mansky and Temin 1995). A taxa de substituição de nucleótidos introduzida pela TR é de aproximadamente  $10^{-4}$  nucleótido por ciclo de replicação, o que é equivalente a um nucleótido substituído por genoma, durante um único ciclo viral (Nowak 1990). A facilidade com que surgem estas mutações origina uma heterogeneidade de estirpes virais, designadas por quasiespécies.

As quasiespécies diferem entre si a nível genotípico e fenotípico, e competem entre si pela sobrevivência e propagação. A subsequente propagação ou domínio de uma estirpe viral em relação a outra é determinado pela sua capacidade de adaptação ao ambiente do hospedeiro, factor particularmente relevante na manutenção de variantes resistentes aos anti-retrovirais. (Fisher *et al.* 1988).

### **1.4 A influência da recombinação genética na diversidade do VIH-1**

Inserções, deleções e duplicações contribuem também para a diversidade genética do vírus. O VIH-1 possui um rápido *turnover* e está estimado que  $10^9$  viriões sejam gerados por dia num indivíduo infectado. Os regimes terapêuticos prolongados e o rápido *turnover* viral, em combinação com altas taxas de mutação, são os primeiros factores que estão na origem do aparecimento de variantes VIH-1, que podem ser resistentes aos anti-retrovirais (Hu and Temin 1990a).



A recombinação genética, entre estirpes do VIH-1, contribui para a sua diversidade (Hu and Temin 1990b). Este facto está associado a possibilidade de estirpes que apresentam diferentes mutações de resistência se poderem recombinar, contribuindo para um aumento significativo das resistências aos fármacos (Wain-Hobson et al. 2003); (Yusa et al. 1997). A recombinação pode ainda levar à aquisição de novas mutações, que compensem a perda de *fitness* viral ou capacidade replicativa, devido à aquisição de mutações anteriores. (Moutouh et al. 1996);(Kellam and Larder 1995)

### **1.5 Pressão selectiva imposta pelos anti-retrovirais.**

A eficácia, de todos os anti-retrovirais actualmente utilizados, está limitada pela emergência de estirpes resistentes, que frequentemente exibem resistências cruzadas às várias classes. (Miller and Larder 2001, Loveday 2001, Loveday 2001) (Deeks 2001).

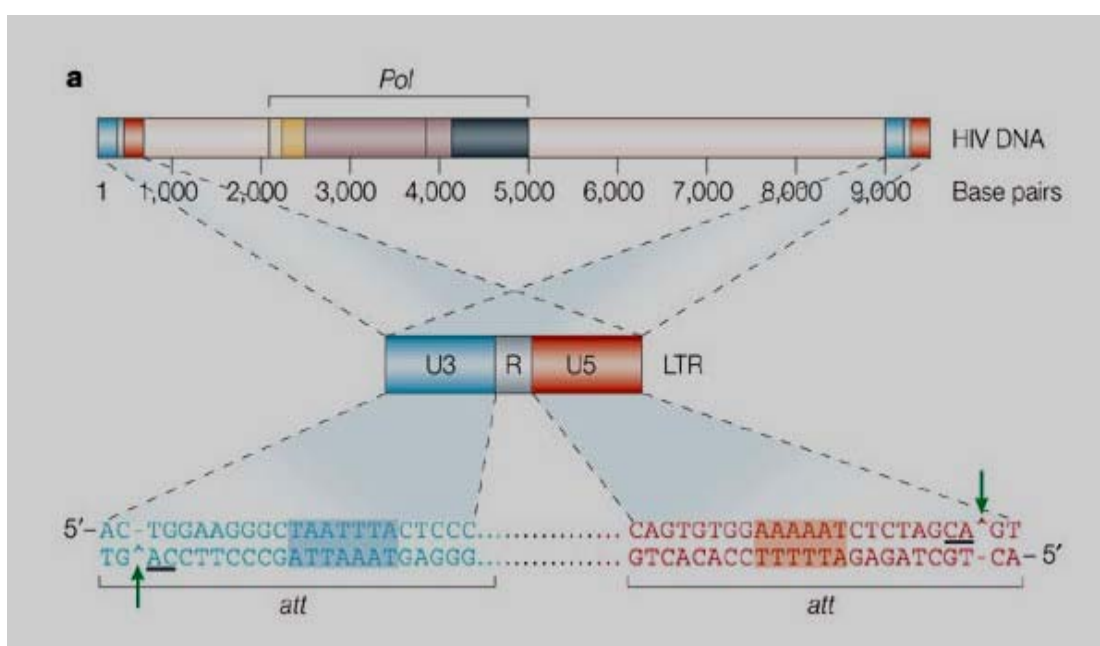
Durante os regimes terapêuticos prolongados, alguns dos viriões das células infectadas poderão sobreviver como variantes resistentes aos fármacos, implicando a subida da carga viral. Na ausência dos anti-retrovirais, as estirpes resistentes apresentam um *fitness* reduzido. Comparativamente a sua congénere *wild-type* (Coffin 1995)

A continuação da pressão selectiva dos anti-retrovirais leva à selecção de mutações secundárias que compensam as mutações primárias, restaurando assim o *fitness* viral para níveis normais.

### **1.6 Estrutura da integrase**

A integrase (p32<sup>IN</sup>) é uma proteína constituída por 288 aminoácidos, com um peso molecular de 32 kD, codificada na extremidade 3' do gene *pol*. Aproximadamente 50 a 100 cópias da enzima integrase são empacotadas por partícula viral. O gene *pol* dos retrovírus

codifica para três enzimas (protease, transcriptase reversa e integrase) que são alvos terapêuticos (Figura 3). Em ambas as extremidades do ADN existem os terminais longos de repetição (LTRs) que consistem em três elementos repetidos (U3-R-U5) com a mesma orientação. A integração requer que os LTRs tenham uma sequência correcta de adeninas e timinas (locais *att*). O processamento 3' catalisa a extracção (setas verdes da figura 2) de dois nucleotídeos (GT) imediatamente após encontrar os di-nucleótidos CA (Figura 2.) gerando terminais 3'-OH reactivos. (Pommier et al. 2005b)



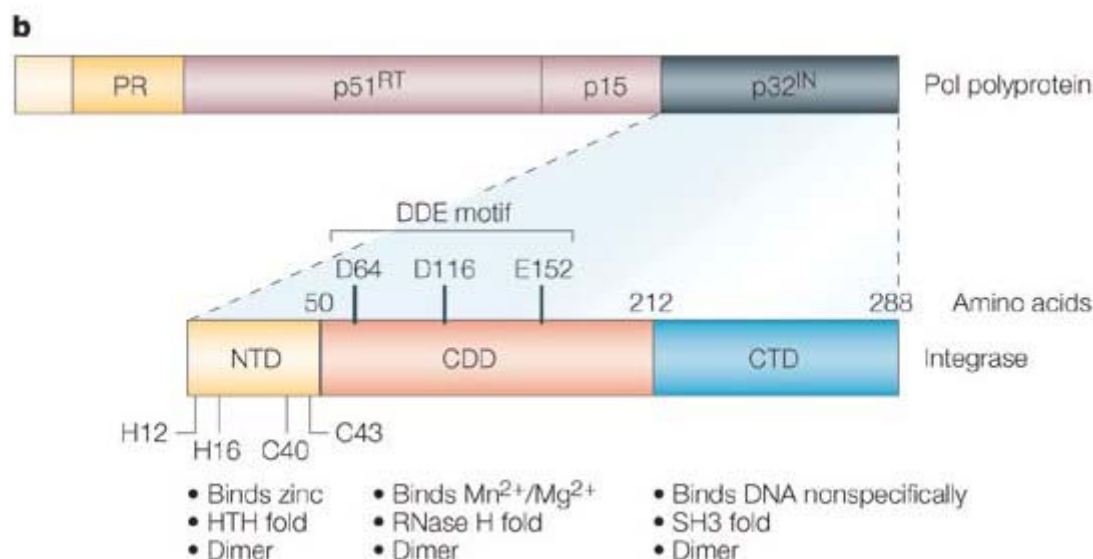
**Figura 2: O gene *pol* do provírus do VIH-1**

Adaptado de: [www.nature.com/.../v4/n3/fig\\_tab/nrd1660\\_F1.html](http://www.nature.com/.../v4/n3/fig_tab/nrd1660_F1.html)

A figura 3 mostra a poliproteína precursora, produto do gene *pol*. A integrase é gerada por clivagem desta poliproteína *pol*, pela protease do VIH-1 durante a maturação viral. A integrase é constituída por três regiões estruturais e funcionais distintas (Pommier et al. 2005b)

- Uma região N-terminal (1-50 aa) necessária à oligomerização da integrase,
- Uma região Central (51-212 aa) responsável pela actividade catalítica,

- Uma região C-terminal (213-288 aa) necessária à ligação do ADN de uma forma não específica.



**Figura 3: As regiões da integrase**

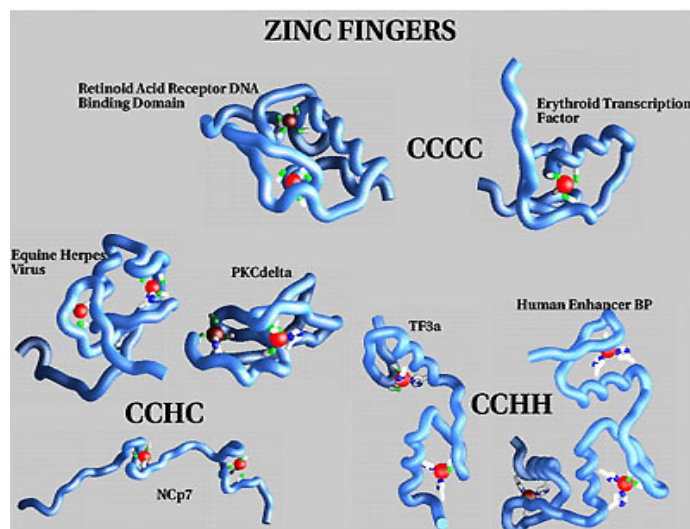
Adaptado de: [www.nature.com/.../v4/n3/fig\\_tab/nrd1660\\_F1.html](http://www.nature.com/.../v4/n3/fig_tab/nrd1660_F1.html)

A estrutura molecular de cada uma destas regiões foi determinada por difração de raio X, ressonância nuclear magnética (RNM) e cristalografia

## 1.7 Região N-terminal

Esta região fica compreendida entre os aminoácidos 1 e 50, e a sua ligação à região Central é absolutamente necessária para o processamento 3'. Esta região interage com o ADN viral apenas quando está ligada à região Central. Esta região N-terminal é caracterizada por possuir dois pares de histidinas e cisteínas (histidina 12 e 16; cisteínas 40 e 43) altamente conservados. Estes dois pares formam um arranjo “HH-CC” chamado dedos de zinco. Os dedos de zinco estão envolvidos na quelação de um átomo de zinco (Cai *et al.* 1997). O zinco estimula a região N-terminal à multimerização da integrase e estabiliza a estrutura terciária

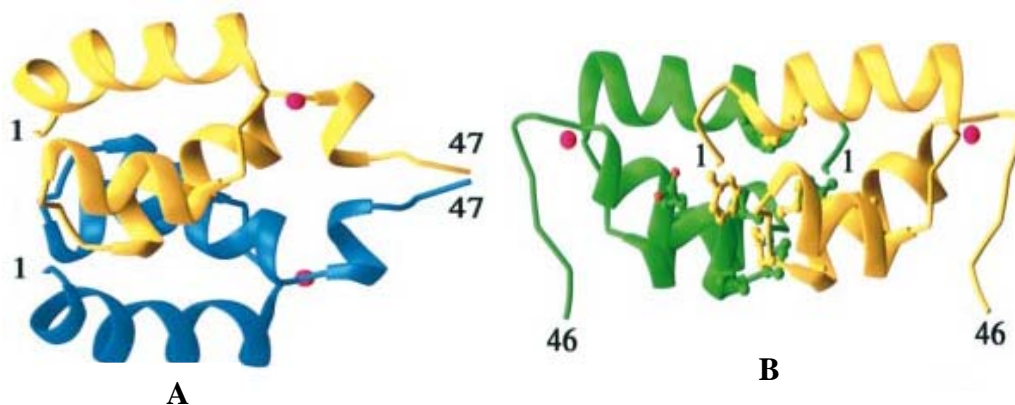
desta, otimizando a actividade catalítica da enzima na ligação ao ADN. A mutação de qualquer um dos quatro aminoácidos do arranjo “HH-CC” reduz a actividade enzimática da integrase (Engelman and Craigie 1992).



**Figura 4: Diferentes, estruturas dedos de zinco.**

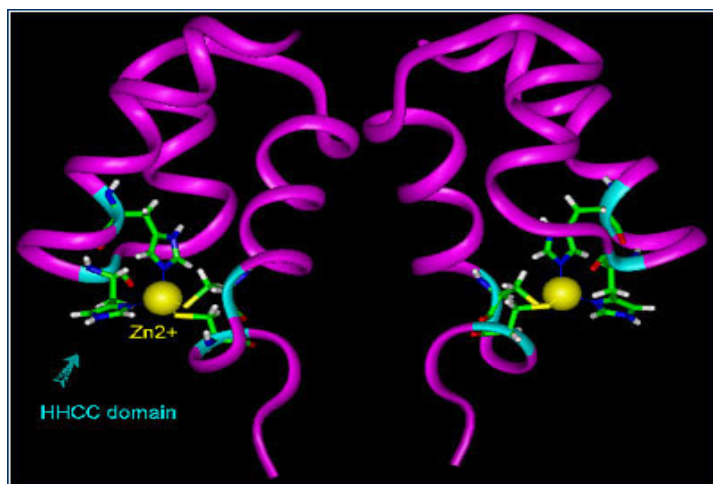
Proteínas com dedos de zinco reconhecem ácidos nucleicos (ADN ou ARN). A figura mostra os diferentes arranjos (HC), que ilustram as distintas estruturas de vários dedos de zinco. O arranjo “CCHC” é típico dos retrovírus. Adaptado de: [www.medgadget.com/archives/img/zinczfm.jpg](http://www.medgadget.com/archives/img/zinczfm.jpg)

Em solução, podemos encontrar esta região sob a forma de monómero (Figura 5a) ou de dímero. É só sob a forma dimérica, que a região N-terminal se liga à região Central (Figura 5b). A sua estrutura dimérica, após ligação à região Central, foi confirmada por resolução RMN. Cada monómero é constituído por quatro hélices  $\alpha$ : as hélices  $\alpha_2$  e  $\alpha_3$  formam um arranjo estável, hélice-volta-hélice, quando complexado com o zinco e pelas interacções hidrofóbicas (Figura6).



**Figura 5: Região N-terminal isolada e ligada a região central**

Análise de interações das regiões. A figura 5 (A) mostra a região N-terminal como uma entidade isolada determinada por RMN. A figura 5 (B) mostra a região N-terminal após ligação com a região central por cristalografia de raios X. Adaptado: [www.nature.com/emboj/journal/v20/n24/abs/7594220a.html](http://www.nature.com/emboj/journal/v20/n24/abs/7594220a.html)



**Figura 6: Ligação do Zinco ao arranjo HHCC**

Adaptado de: [www.msg.ucsf.edu/stroud/structure/hiv\\_int.htm](http://www.msg.ucsf.edu/stroud/structure/hiv_int.htm)

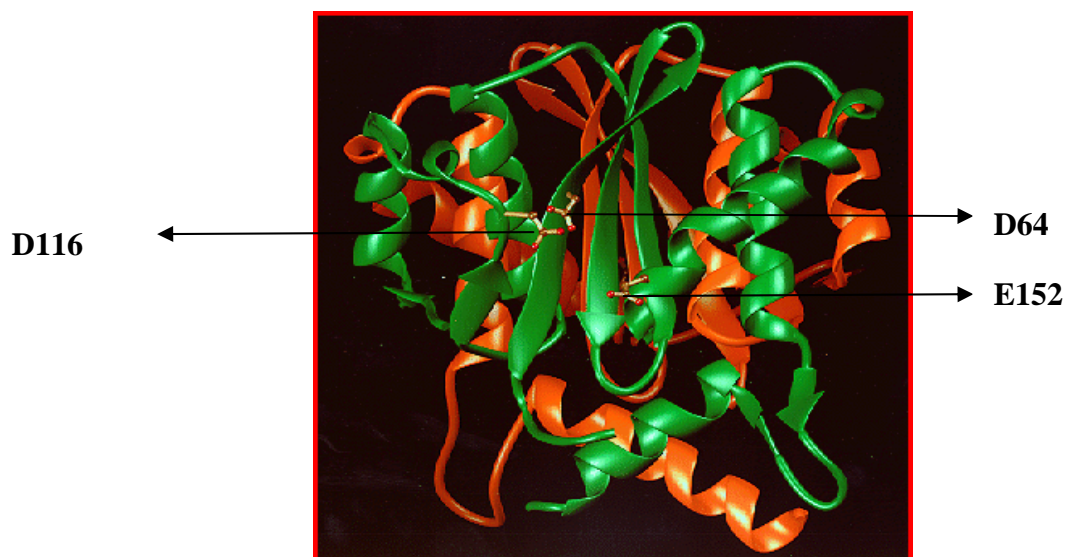
Durante a formação do complexo de pré-integração (CPI), a região N-terminal é a região preferencial para a ligação de dois factores de transcrição celular (INI1 e LEDGF/p75) (Kalpana et al. 1994b), (Cherepanov et al. 2003).

## 1.8 Região Central

A região central fica compreendida entre os aminoácidos da posição 51 e 212, e apresenta-se sob uma forma dimérica, tanto em solução como na estrutura da integrase. O enrolamento da estrutura da região central é semelhante, ao encontrado na estrutura enrolada da RNase H, região catalítica da RuvC e MuA transposase (Jegade *et al.* 2008). Estas enzimas pertencem à classe enzimática das ADN-transferases de polinucleótidos na qual se inclui a integrase. A actividade enzimática destas enzimas consiste na clivagem e junção de ADN por trans-esterificação. (Dyda *et al.* 1994)

O centro catalítico ou tríade catalítica caracteriza a região central da integrase e é constituído sempre por três aminoácidos invariáveis (Figura7):

- ácido aspártico D64,
- ácido aspártico D116,
- ácido glutâmico E152.



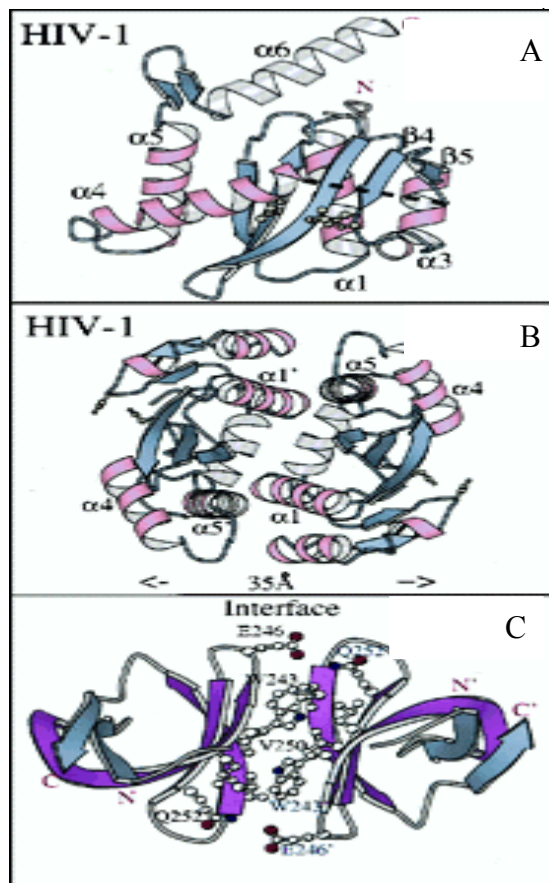
**Figura 7: Estrutura dimérica da região Central**

A ilustração mostra a estrutura dimérica, na qual os aminoácidos aspartato e glutamato (D64, D116 e E152) formam uma tríade catalítica. A estrutura tridimensional tem características em comum com RNase H e outras transferases de polinucleotídeos. Adaptado de: [www.gtmb.org/VOL3A/GTMBVOL3AHTML/28\\_Faust.htm](http://www.gtmb.org/VOL3A/GTMBVOL3AHTML/28_Faust.htm)

O centro catalítico estabelece ligações com o  $Mg^{2+}$  ou  $Mn^{2+}$  (Khan *et al.* 1991, Kulkosky *et al.* 1992). A tríade catalítica é essencial a todas as funções catalíticas da integrase e encontra-se conservada em todas as integrases de retrovírus, bem como as retrotransposases de plantas, animais e fungos e ainda em algumas transposases bacterianas (Rice *et al.* 1996)

### **1.8.1 Características da Região Central:**

A figura 8A mostra as folhas de um modelo da região central da integrase do VIH-1 (Bujacz *et al.* 1996). A tipologia da região central do vírus é muito semelhante à de outras integrases, nomeadamente a do vírus do sarcoma aviário -VSA.



**Figura 8: Organização da região Central**

Adaptado de : <http://www.jbc.org/content/271/33/19633.full>

A região Central é formada por 5 folhas  $\beta$ , das quais três são anti-paralelas ( $\beta 1, \beta 2$ , e  $\beta 3$ ) e que se encontram rodeadas por 6 hélice  $\alpha$ . O contacto inicial entre os monómeros da região Central para formarem um dímero, envolve o contacto da hélice  $\alpha 1$  de um monómero com a hélice  $\alpha 5$  de outro monómero (Figura 8B). A dimerização de centro catalítico é fundamental para a multimerização exigida para actividade catalítica da integrase (Andrake and Skalka 1995);(Jenkins et al. 1996).

Para além do centro catalítico, a região central também é caracterizada por possuir outros domínios funcionais e aminoácidos importantes.



- Sinalizadores de localização nucleares (NLS), sequências críticas na mediação da importação nuclear da integrase.(Bouyac-Bertoia et al. 2001b).
- A sequência K<sub>186</sub>R<sub>187</sub>K<sub>188</sub> para a multimerização na interface dímero:dímero (Berthoux *et al.* 2007)
- Os aminoácidos:  
(H12,L102,A128,A129,C130,W131,W132,I161,R166,Q168,E170) - estão envolvidos em ligações químicas e contactos hidrofóbicos com o factor de crescimento *lens epithelium* (LEDGF/p/75). O LEDGF/p/75 é um co-factor celular usado para integração do VIH-1, ligando a integrase à cromatina. (Busschots *et al.* 2007)
- Os aminoácidos 140-149 formam um laço dito “catalítico”. A flexibilidade deste laço permite que o aminoácido D152 se oriente na direcção dos outros dois aminoácidos da tríade catalítica, D64 e D116 (Figura 9). A estrutura exacta do laço não é totalmente conhecida mas sabe-se que varia de acordo com determinadas mutações. A mutação G140A/G149A induz uma perda de flexibilidade do laço, traduzindo-se numa diminuição drástica da actividade enzimática, embora sem afectar a interacção com o ADN. Isto sugere que a flexibilidade conformacional do laço é essencial para a etapa catalítica pós fixação ao ADN viral. Esta necessidade é confirmada pelo estreito contacto entre o aminoácido G148 e a extremidade do ADN viral após o processamento 3’ (Johnson et al. 2006). A flexibilidade parece primordial para poder acomodar potenciais inibidores, como sugerem os ensaios que combinam dinâmica molecular e “*docking*” de inibidores (Carlson *et al.* 2000).

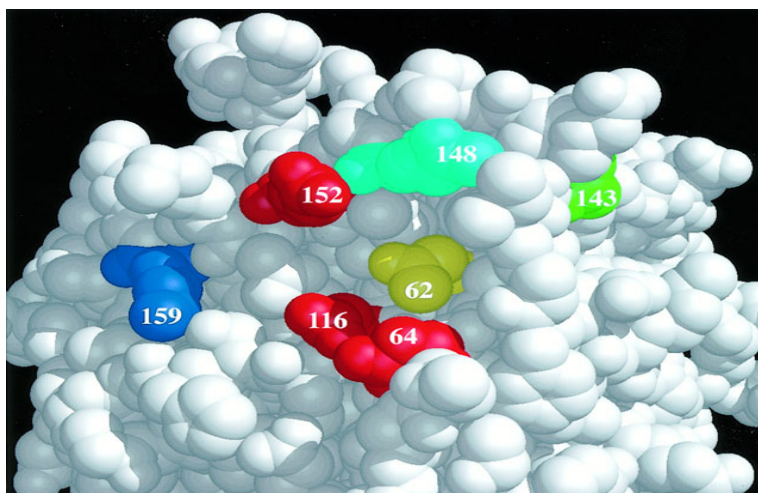
- Os aminoácidos importantes na interação com o ADN viral encontram-se localizados na proximidade do centro catalítico da região central e são: Q148,Y143,K156,K159. A mutação de dois deles, a lisina 156 e 159, determina o fim da especificidade ligação (Figura10).

A região Central tem um papel específico no correcto posicionamento da integrase no ADN viral. (Harper *et al.* 2003)



**Figura 9:Laço flexível do Centro catalítico**

Adaptado de : [www.msg.ucsf.edu/stroud/img/figures/Int\\_core.jpg](http://www.msg.ucsf.edu/stroud/img/figures/Int_core.jpg)



**Figura 10: Estrutura da região central na proximidade da tríade catalítica da integrase**

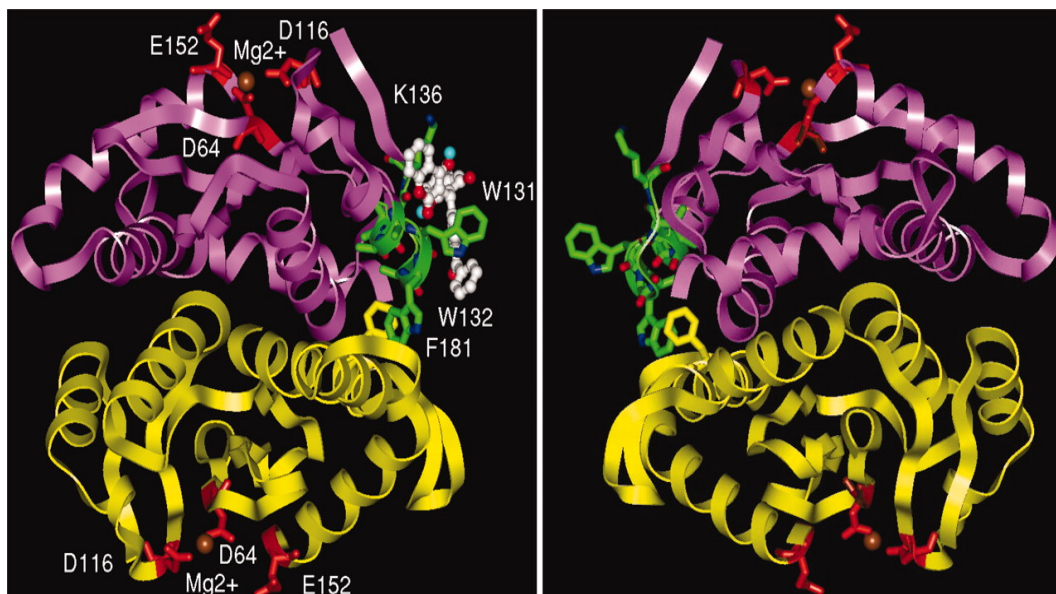
A estrutura da integrase que envolve o centro activo é mostrada a branco. Os três aminoácidos do centro catalítico, D64, D116 e E152 são identificados a vermelho, enquanto o K159 é mostrado a azul. Y143 é mostrada a verde, e o Q148 em ciano. Adaptado de : [www.nature.com/.../v17/n19/full/7591277a.html](http://www.nature.com/.../v17/n19/full/7591277a.html)

## 1.9 Fixação do $Mg^{2+}$ ao centro catalítico

A cristalografia da região central revela que o sal que se liga ao centro catalítico é o  $Mg^{2+}$ . (Dyda et al. 1994) (**Erro! A origem da referência não foi encontrada.**).

A ligação é estabelecida entre o oxigénio dos grupos carboxílicos do D64 e D116 e quatro moléculas de água, numa geometria octaédrica. Experiências de dinâmica molecular sugerem a possível existência de um segundo ião  $Mg^{2+}$  ligado ao centro catalítico. (Lins *et al.* 2000).

Sabe-se que a fixação do magnésio induz uma alteração da conformação do centro catalítico, que difere segundo a presença ou ausência do ADN viral (Asante-Appiah *et al.* 1998). Esta incerteza na determinação da estrutura exacta do centro catalítico e a possível existência de um segundo ião magnésio dificulta o desenho ou a melhoria dos inibidores da integrase existentes. Os inibidores disponíveis são concebidos para actuarem como quelantes do magnésio.

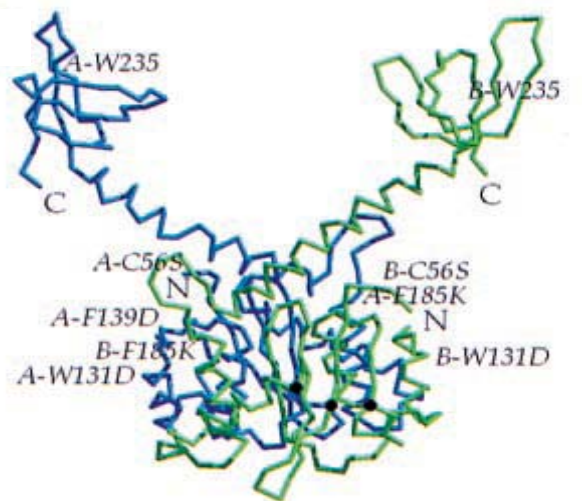


**Figura 11: Ligação do Mg<sup>2+</sup> à tríade catalítica**

Adaptado de: <http://www.pnas.org/content/103/26/10080/F4>.

## 1.10 Região C-Terminal

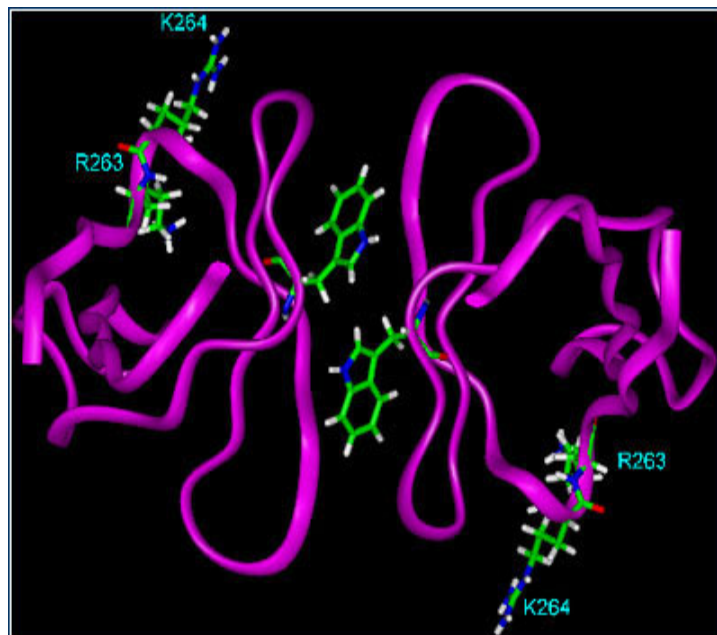
Esta região fica compreendida entre os aminoácidos 213 e 288 e é a região menos conservada da integrase. A sua função principal é estabilizar as ligações da integrase ao ADN, embora esta interação não seja específica. A estrutura obtida por RMN mostra que a região C-terminal, também forma dímeros quando em solução. A sua forma monomérica estabelece ligações com a região Central dimérica. Os monómeros estão orientados paralelamente, sendo cada um constituído por 5 folhas  $\beta$  anti-paralelas, que se apresentam numa estrutura dita em barril (figura 13).



**Figura 12: Ligação da região Central à região C-terminal**

Adaptado de: [www.msg.ucsf.edu/.../chen\\_PNAS\\_2000.htm](http://www.msg.ucsf.edu/.../chen_PNAS_2000.htm)

A região C-terminal liga-se à TR (transcriptase reversa) e esta interacção é necessária à própria actividade da TR (Zhu *et al.* 2004).



**Figura 13: Ligação da região C-terminal através do domínio SH3 ao ADN**

O domínio SH<sub>3</sub> dobra a região C-terminal, estabiliza a ligação ao ADN através dos aminoácidos R263 e K264. O domínio SH<sub>3</sub> clássico é geralmente encontrado em proteínas que interagem com outras proteínas e medeiam a montagem de complexos de proteicos específicos. Adaptado de: [www.msg.ucsf.edu/stroud/structure/hiv\\_int.htm](http://www.msg.ucsf.edu/stroud/structure/hiv_int.htm)

### 1.11 Actividade catalítica da integrase

A integração do ADN do VIH-1 em qualquer zona do ADN da célula hospedeira é fundamental para a continuação da infecção. (Blankson *et al.* 2002). Após a transcrição reversa do ARN genómico do VIH-1, o ADN de cadeia dupla resultante serve de substrato a duas reacções catalisadas pela integrase.

No citoplasma a integrase liga-se a uma sequência específica, presente nas longas sequências terminais de repetição (LTRs) do ADN viral (Pommier *et al.* 2005b), catalisando a remoção do di-nucleotido GT imediatamente após encontrar os di-nucleotidos CA, no sentido 3', em ambas as cadeias de ADN viral. Esta remoção é feita através de um ataque nucleofílico as ligações fosfodiésteres entre a deoxiguanosina e a deoxiadenosina (Figura 14). O passo descrito consiste no processamento 3' no qual a integrase se mantém ligada ao LTR, formando o complexo de pré-integração (CPI).

O CPI contém:

- Proteínas virais (matriz, Vpr, p7/nucleocapside, transcriptase reversa) (Bukrinsky *et al.* 1992)
- ADN viral,
- proteínas do hospedeiro (a interactor-1 (INI1) (Kalpana *et al.* 1994a),
- factor de crescimento *lens epithelium* (LEDGF/p75) (Cherepanov *et al.* 2003),
- proteína 60 do choque térmico (HSP60) (Parissi *et al.* 2001).

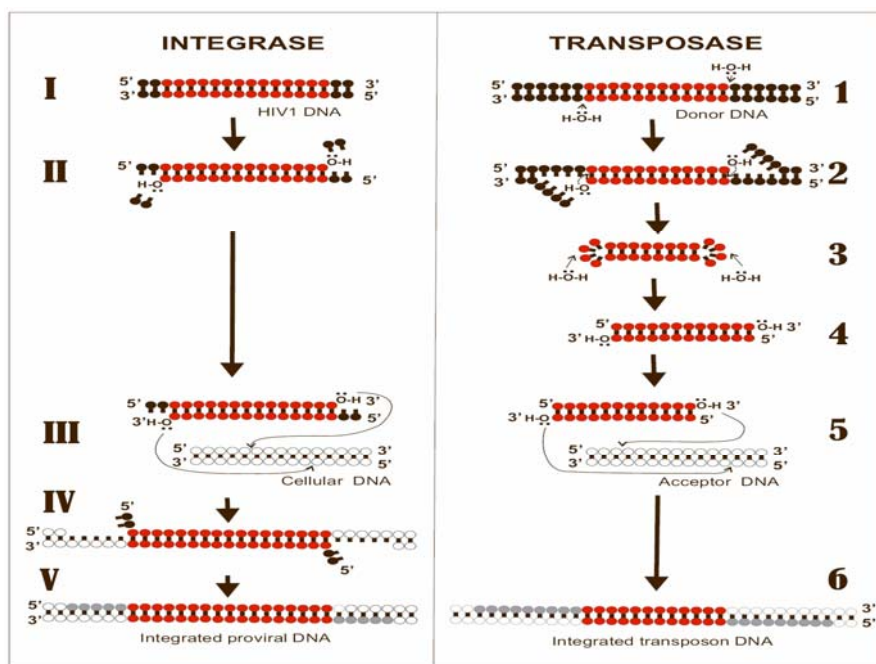
O CPI é transportado de forma activa para o núcleo, onde ocorre a segunda reacção designada por junção 3' ou transferência de cadeias.

Após a transferência do CPI para o interior do núcleo ocorre a ligação do ADN viral ao ADN do hospedeiro. Esta ligação tem por base as propriedades nucleofílicas do CPI para se ligar ao ADN do hospedeiro através das extremidades 3'-OH do ADN viral. Ambas as extremidades 3'-OH são mantidas muito próximas, e são integradas na extremidade 5' do ADN hospedeiro separadas por um intervalo de 5 pares de bases, característico do VIH-1. (Johnson *et al.* 2004).

Os dois nucleótidos não emparelhados no terminal 5' do ADN viral são removidos e o espaço gerado no local da integração é preenchido, provavelmente, por enzimas de reparação celular, uma vez que *in vitro* estes espaços não são preenchidos, mesmo na presença do CIP. (Yoder and Bushman 2000). A junção 3' permite completar a integração do ADN viral no genoma do hospedeiro.

O ADN viral integrado é agora designado como provirus (ADN proviral) e torna-se no molde para a síntese de novos viriões, processo que será executado pelos processos celulares.

A estrutura e a actividade biológica da integrase confirmam que é uma enzima que pertence ao grupo bioquímico das nucleases/transferases de polinucleotidos (Grindley and Leschziner 1995);(Rice et al. 1996)



**Figura 14: Sequência de eventos da integrase e da transposase**

Sequência de eventos da integrase (esquerda) e Tn5 transposase (direita) na integração de ADN.  
Adaptado de: *Retrovirology* 2007 4:21:

### 1.12 A importância do complexo de pré-integração

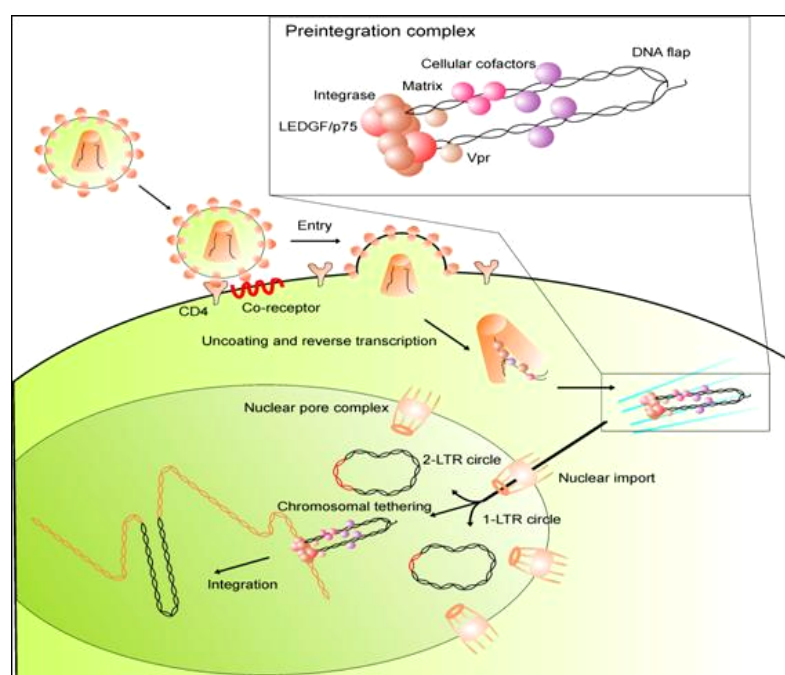
A integrase possui sequências que funcionam como sinais de localização nuclear (SLN) e pelo menos uma dessas sequências encontra-se localizada na região C-terminal, interagindo com um receptor de importação nuclear (Ao *et al.* 2007).

A sequência constituída pelos aminoácidos V165 e R166, que se localiza na proximidade do centro catalítico contribui para aumentar a afinidade da integrase pelo ADN nuclear. A mutação destes aminoácidos não anula completamente a translocação dos CIPs, sugerindo a presença de mecanismos complementares ou redundantes de translocação.

O CPI do VIH-1 é transportado no citoplasma através do sistema microtubular, atravessando os poros nucleares (McDonald *et al.* 2002). A afinidade do CPI para o ADN do núcleo da célula hospedeira, permite ao VIH-1 replicar-se em células não proliferativas tais como os macrófagos (Bukrinsky *et al.* 1992).



A proteína A1HMGA1, conhecida como HMG1(Y) e o factor de barreira à auto integração (BAF), regulam directamente a ligação da integrase ao ADN celular. A HMGA1 estimula a actividade da integrase (Hindmarsh et al. 1999);(Farnet and Bushman 1997) e o BAF estimula a integração e inibe a auto integração, interagindo com a emerina. A emerina, transportador localizado nos poros nucleares, liga a membrana nuclear interna à cromatina (Jacque and Stevenson 2006);(Shun et al. 2007).



**Figura 15:Complexo de pré-integração**

O VIH depende da interacção com várias proteínas celulares para completar o seu ciclo de replicação. A figura mostra a importação nuclear e a integração. Adaptado de : [www.kuleuven.be/.../research\\_aids\\_research.html](http://www.kuleuven.be/.../research_aids_research.html)

Na ausência de integração, o ADN viral linear é circularizado, originando um ADN circular sem capacidade replicativa. Estas formas circulares permanecem no núcleo durante, apenas um certo período de tempo, após o qual são eliminadas (Lewinski and Bushman 2005). Este facto, sugere que uma das estratégias terapêuticas que poderá ser utilizada será a utilização de inibidores de transferência de cadeias, que conduzirão à acumulação destas

formas circulares. Após o processo de integração estar completado, a enzima é degradada nos proteossomas. (Mousnier *et al.* 2007)

### 1.13 A importância do $Mg^{2+}$ nas reacções catalíticas da integrase

O processamento 3' e a transferência de cadeias requerem um co-factor metálico. Poderá tratar-se do  $Mg^{2+}$  ou do  $Mn^{2+}$ . No entanto, *in vivo*, o co-factor mais provável será o  $Mg^{2+}$ . Numerosos dados sugerem, que as reacções dependentes do  $Mg^{2+}$  são as mais representativas de actividades fisiológicas. Em contrapartida, a integrase possui uma forte actividade, nuclease não específica, em presença do  $Mn^{2+}$  (Engelman and Craigie 1995). Acresce ainda o facto de que a tolerância a variações de nucleótidos nas extremidades do ADN viral ser muito maior na presença do  $Mn^{2+}$  do que na presença do  $Mg^{2+}$  (Esposito and Craigie 1998). Do ponto de vista enzimático, numerosas mutações da integrase, permanecem silenciosas nos testes dependentes do  $Mn^{2+}$  mas não nos testes dependentes do  $Mg^{2+}$ . Por exemplo, *in vivo*, mutações de aminoácidos no arranjo HHCC afectam o processamento 3' e a integração nos testes que utilizam o  $Mg^{2+}$  mas o mesmo já não acontece se forem realizados na presença de  $Mn^{2+}$ . O estímulo do zinco, na actividade da integrase, não depende da presença do  $Mg^{2+}$  como co-factor (Lee and Han 1996). A diferença entre os co-factores utilizados tem repercussões em termos farmacológicos, uma vez que alguns fármacos inibidores da actividade da integrase, não têm a mesma eficácia na presença de  $Mg^{2+}$  ou de  $Mn^{2+}$ .

Actualmente ainda não foi possível determinar a estrutura completa da integrase por cristalografia, mas o conhecimento da forma como as suas subunidades se ligam para formarem dímeros evoluiu muito nesta última década. A forma como a região N-terminal se liga à região Central, que contém o centro catalítico, foi fundamental no desenho de novos inibidores da integrase - Raltegravir e Elvitegravir. As ligações entre os aminoácidos das diferentes regiões da integrase contribuem para diferentes tipos de conformações da integrase.

A determinação dos aminoácidos intervenientes nestas ligações é fundamental, para valorizar os aminoácidos mutados e compreender os polimorfismos naturais da integrase.

## 2 Capítulo II - Inibidores da integrase

### 2.1 Introdução.

Durante muitos anos foi consensual que um anti-retroviral dirigido contra a integrase, seria um complemento importante na terapêutica *HAART*. No entanto, existiam algumas dúvidas sobre a capacidade destes inibirem a replicação viral *in vivo*. Por exemplo, quando uma célula é infectada pelo vírus são introduzidas na célula numerosas cópias da enzima integrase, mas apenas uma é necessária para a realização de dois eventos de integração que conduzem à formação do provirus. Assim a integrase, foi interpretada como sendo um alvo difícil de inibir. Tal dúvida sabe-se hoje ser infundada com a identificação de compostos que de uma forma inequívoca inibem a replicação do HIV-1 através (Johnson et al. 2004); (De Clercq 2002) da inibição da integrase. Estes inibidores demonstram uma maior afinidade para integrase na presença do ADN, do que para a integrase purificada e isolada. (Pommier et al. 2005a)

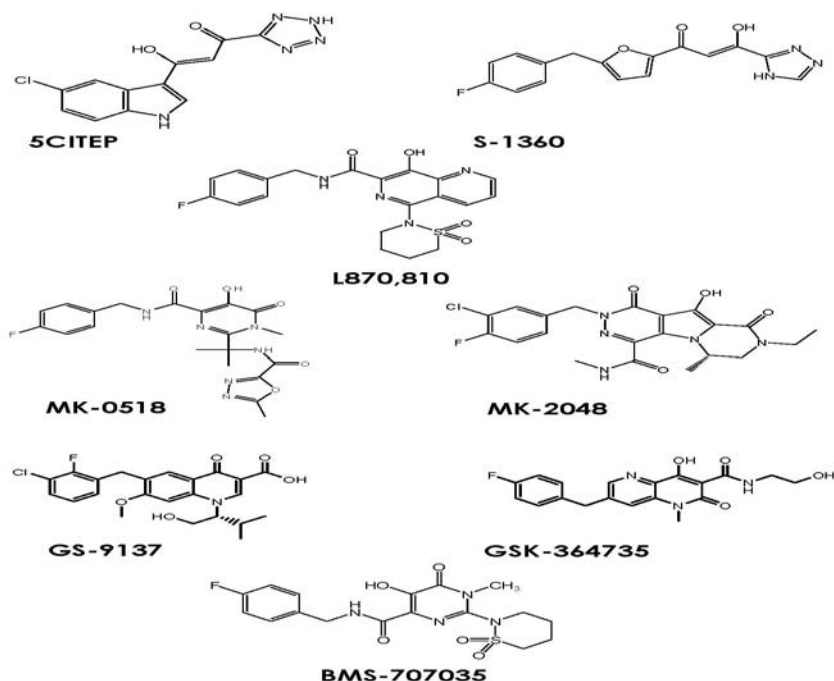
### 2.2 Mecanismo de acção dos inibidores

Ao contrário dos inibidores competitivos que competem simplesmente pelo substrato, os inibidores da integrase bloqueiam a reacção de transferência de cadeias. Os inibidores da integrase são designados de inibidores catalíticos. No geral estes inibidores podem ser classificados em dois grupos:

- Inibidores do processamento 3' e transferência de cadeias,

- Inibidores da transferência de cadeias.

A ligação da integrase ao ADN viral induz alterações conformacionais da mesma (Smolov *et al.* 2006), permitindo desta forma que os inibidores se possam ligar a regiões distintas da integrase. Os inibidores interagem como centro catalítico, inibindo a funcionalidade enzimática da enzima. Por exemplo, os derivados da quinolina só inibem o processamento 3' (Zouhiri *et al.* 2000). Outros ácidos  $\beta$ -dicetos (inibidores) contêm compostos que inibem ambas as reacções: os protótipos L-708, 906 e L-731, 988 inibem a transferência de cadeias. Composto 5CITEP contém o ácido diceto (Figura 16) que inibem ambas as reacções. Os inibidores com as duas funcionalidades podem ligar-se ao ADN viral e celular. Os inibidores de transferência de cadeias ligam-se de forma selectiva nos locais próprios de ligação do ADN celular. (Marchand *et al.* 2002a). Mas mais importante ainda, é que os inibidores da reacção de transferência de cadeia reconhecem a conformação dos locais activos da integrase. Estes locais activos só ficam definidos após a ligação aos LTRs do ADN viral. As potencialidades deste tipo de inibição ainda não são totalmente compreendidas. Pode-se adiantar que os inibidores da reacção de transferência de cadeia demonstram uma maior selectividade e especificidade para integrase viral do que os inibidores do processamento 3'.

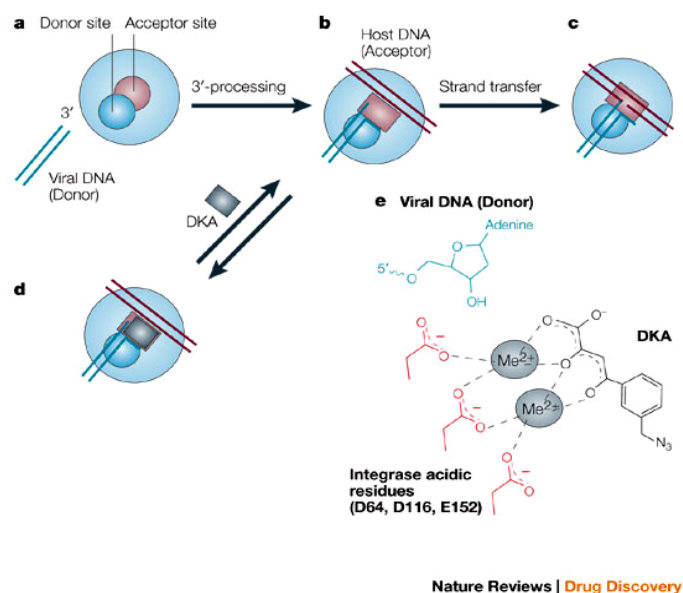


**Figura 16: Estrutura molecular de (DKA) ácidos diceto-aril**

Adaptado de : Retrovirology 2009 6:25:10.1186/1742-4690-6-25

No modelo da figura 17 o DKA (ácido diceto-aril) inibe a integrase na interface entre o ADN viral e o  $Mg^{2+}$ . O DKA inibe de forma selectiva a transferência, ligando-se na interface do complexo integrase/ADN. A integrase possui dois locais de ligação: um local de ligação para o ADN viral (local dador) e um local de ligação para o ADN celular (local receptor). A formação do complexo integrase/ADN induz uma alteração conformacional do local receptor, tornando-o competente para se ligar ao ADN celular. O que acontece normalmente é a ligação da integrase ao ADN celular, seguindo-se a transferência de cadeias do ADN viral. Na presença de DKAs este liga-se ao local receptor, mas só após ter ocorrido o processamento 3'. Na interface do complexo formado pela integrase/ADN/ $Mg^{2+}$ , o ADN viral processado e ligado à integrase ataca as ligações fosfodiésteres do ADN hospedeiro. É proposto, que os DKAs quelatam o  $Mg^{2+}$  do centro catalítico e estabilizam o complexo integrase/ADN na etapa do processamento 3'. Os DKAs pertencem a um grupo emergente de inibidores, designados por

inibidores de interface, uma nova classe de fármacos que possui um modo único de acção. (Espeseth et al. 2000) (Marchand et al. 2002b).

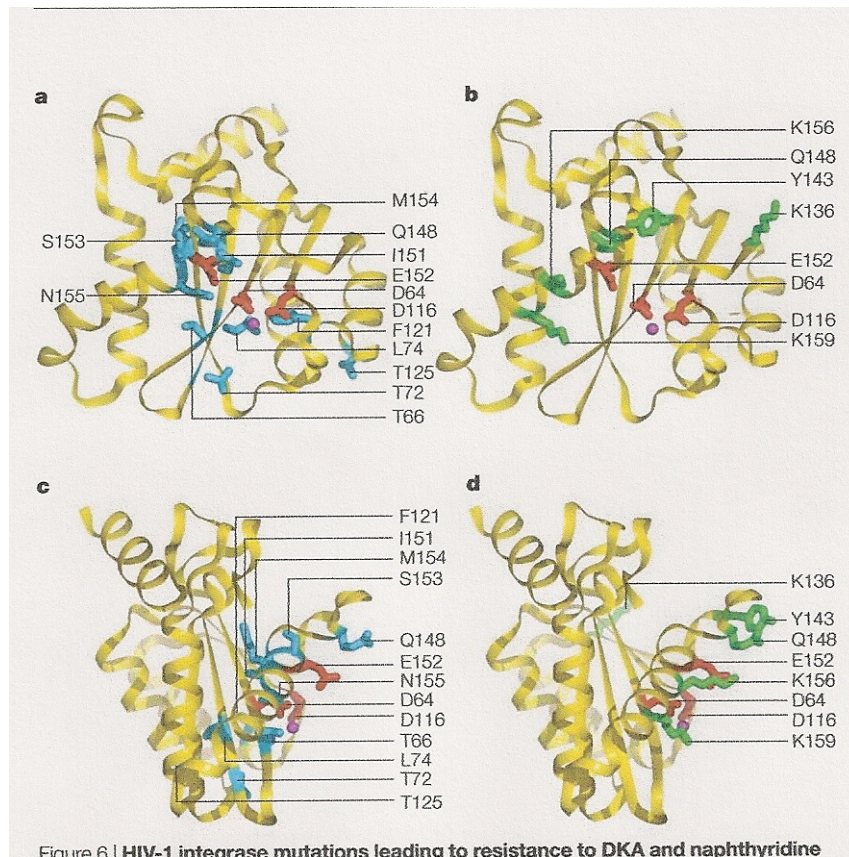


**Figura 17: Mecanismo de ligação dos inibidores à integrase**

Adaptado: [www.nature.com/.../v4/n3/fig\\_tab/nrd1660\\_F1.html](http://www.nature.com/.../v4/n3/fig_tab/nrd1660_F1.html)

De acordo com este modelo, é fundamental que a tríade catalítica esteja complexada pelo Mg<sup>2+</sup>, de forma a permitir a ligação do DKA ao local receptor da integrase.

Verificou-se que integrases mutantes e cataliticamente inactivas falham a ligação ao DKA (Grobler *et al.* 2002). A ligação do DKA, também requer a presença conservada das extremidades do ADN viral, uma vez que o ADN viral com extremidades alteradas, não permite essa ligação. (Espeseth et al. 2000).



**Figura 18: Aminoácidos envolvidos na resistência aos DKA**

Adaptado: [www.nature.com/.../v4/n3/fig\\_tab/nrd1660\\_F1.html](http://www.nature.com/.../v4/n3/fig_tab/nrd1660_F1.html)

As mutações seleccionadas pelo DKAs podem-se situar na região Central (Hazuda *et al.* 2000), e na região C-terminal (Fikkert *et al.* 2003b). Os aminoácidos da região Central, que conferem resistências ao DKA, estão situados em torno da tríade catalítica. Este grupo de aminoácidos sobressai pela sua importância na ligação ao ADN. Por exemplo, o aminoácido Q148 parece estar envolvido no alinhamento do ADN viral antes da transferência de cadeias para o ADN celular. ((Fikkert *et al.* 2003a)

### 3 Capítulo III - Polimorfismos da integrase

#### 3.1 Introdução

Os inibidores da integrase representam os mais recentes anti-retrovirais associados ao tratamento de doentes infectados com o VIH-1.

Actualmente estão identificadas mais de 60 mutações no gene da integrase, associadas à resistência tanto *in vitro* como *in vivo* aos diferentes inibidores da integrase, 40 dessas mutações estão associadas ao Raltegravir e ou Elvitegravir (Ceccherini-Silberstein et al. 2009a), fármacos já disponíveis comercialmente ou em fase III de ensaios clínicos.

Os polimorfismos que ocorrem naturalmente no gene da integrase podem ter implicações importantes no desenvolvimento de resistência aos inibidores da integrase. A eficácia destes inibidores e as vias que levam à resistência, no grupo Major do VIH-1 continuam em estudo. Na maioria das análises de sequências do grupo M, os polimorfismos são definidos como mutações presentes em 0,5% ou mais das sequências analisadas (Rhee *et al.* 2008). A variabilidade de um aminoácido numa determinada posição é calculada através da percentagem de sequências que contêm esse aminoácido no lugar do aminoácido mais comum para essa posição (Colson *et al.* 2005). Neste estudo, foram analisados os polimorfismos presentes no gene da integrase de estirpes VIH-1, pertencentes ao grupo M. A amostragem é constituída por 126 amostras pertencentes a indivíduos infectados pelo VIH-1 sem terapêutica anti-retroviral ou que nunca fizeram terapêutica com inibidores da integrase.



## **3.2 Objectivos**

1. Caracterização dos aminoácidos da integrase do VIH-1.
2. Identificação dos polimorfismos da integrase.
3. Identificação das posições muito variáveis (>10%)

## **3.2 Material e métodos**

### **3.2.1 Amostras**

As amostras para a realização do estudo foram obtidas por punção venosa, colhidas em tubos de EDTA e centrifugadas a 2500 g durante 10 min. O plasma foi separado e conservado a -80°C até à realização do teste.

### **3.2.2 População alvo**

A amostragem para estudo foi retirada da população de doentes VIH-1, do laboratório de biologia molecular do serviço de Medicina Transfusional do Centro Hospitalar Lisboa Ocidental, Hospital Egas Moniz. Foram seleccionados doentes que nunca fizeram terapêutica anti-retroviral ou a fazer terapêutica anti-retroviral mas sem inibidores da integrase. Foram estudadas 126 amostras com carga viral detectável (> a 500 cópias/ml). A recolha das amostras e realização dos testes decorreu entre Junho e Novembro de 2009.

### 3.2.3 Extracção do ARN viral

Para a extracção do ARN viral do HIV-1 foi utilizado o equipamento automático Nuclisense® easyMag®, da casa comercial Biomerieux, de acordo com o protocolo fornecido. A extracção de ácidos baseia-se na utilização de sílica e partículas magnéticas.

### 3.2.4 Protocolo de amplificação

Após a extracção do ARN viral, é efectuada uma transcrição reversa do ARN viral para ADN complementar, seguida de uma *Polimerase Chain Reaction* (RT-PCR). O protocolo de amplificação utilizado neste estudo é fornecido no *kit* comercializado pela Celera com a designação de Viroseq™ HIV-1 Integrase Genotyping Kit. A amplificação incluiu um controlo positivo HIV-1 com uma concentração inicial de 800000 cópias/ml que foi diluído a 1:10 e um controlo negativo (plasma). A preparação da RT-PCR *master mix* realizou-se numa área de pré-amplificação (tabela1).

Tabela 1: *Master mix* para uma amostra

Reagentes	Volume para uma reacção (µl)
RT-PCR Submix	36.0
RT-PCR enzyme mix	2.0
UNG, heat-labil	1.0
RNase Inhibitor	1.0
Volume final	40.0

*Master mix:*

- RT-PCR Submix: Tris/KCL/(  $\text{NH}_4$  )<sub>2</sub>SO<sub>4</sub> tampão com DDT e MgCl<sub>2</sub>, dATP, dCTP, dGTP, dTTP < 0,1%; oligonucleótidos sintéticos não infecciosos de VIH-1.
- UNG, Heat-Labil:  $\geq 1\text{U}/\mu\text{L}$  Uracil-DNA Glicosilase termolábil em 50% glicerol.

- RNase Inhibitor, 20 U/μL RNase inibidor em 50% de glicerol.
- RT-PCR enzyme mix: transcriptase reversa *omniscrypt* e *sensiscrypt*, taq DNA polimerase HotStar.

Manter as enzimas no gelo durante a preparação da *master mix*. Adicionar 40 μl da *master mix* à temperatura ambiente a cada poço da placa de reacção ou tubo de reacção. Adicionar 10 μl do ARN extraído a cada tubo. Fechar os tubos. Homogeneizar, e efectuar uma breve centrifugação. Colocar a placa no termociclador numa área de amplificação. Iniciar o programa de amplificação descrito na tabela 2

Tabela 2: Ciclos de amplificação da RT-PCR da integrase

Fase	Temperatura °C	Tempo (min)	Número de ciclos
UNG	20	10:00	-
Transcrição reversa	55	30:00	-
Activação da PCR	95	15:00	-
Desnaturação	94	00:15	45
Emparelhamento	55	00:30	
Extensão	68	01:30	
Extensão	68	7:00	-
Manter	4	Manter	-

### 3.2.5 Quantificação dos produtos da RT-PCR

Preparação do gel de agarose a 1% usando Tris/Borato/EDTA 1X (1gr de agarose em 100ml de TBE). Aquecer 2 min no micro-ondas a 600 watts, depois de arrefecer adicionar 1µl de brometo de Etídio. Colocar o gel numa cama de electroforese e deixar arrefecer, até à sua utilização na tina de electroferese. Colocar o marcador de peso molecular ADN e o tampão “*Loading Buffer*” à temperatura ambiente.

Colocar na primeira e na segunda posição do gel, respectivamente 6 µl e 3 µl de Marcador de peso molecular DNA. Nas restantes posições são colocados 10 µl de cada amostra, resultante da mistura de 5 µl de *Loading Buffer* (bromofenol) com 5 µl de produto de RT-PCR. Efectuar a electroferese a 10V/cm até que o bromofenol tenha migrado pelo menos 5cm no gel de agarose. Avaliar o gel com luz ultra-violeta. Fotografar com um tempo de exposição que permita diferenciar as bandas do marcador e as bandas dos produtos de RT-PCR. Avaliar a quantidade dos produtos de RT-PCR em cada amostra por comparação da intensidade de cada banda com a intensidade das bandas do marcador, tendo como referência a tabela 3.

Tabela 3: Tabela de referência para a quantificação dos produtos de RT-PCR

<b>Bandas do marcador de peso molecular</b>	<b>Tamanho (Kb)</b>	<b>Linha 6µl (ng)</b>	<b>Linha 3µl (ng)</b>
Primeira	2,0	100	50
Segunda	1,2	60	30
Terceira	0,8	40	20

### 3.2.6 Purificação dos produtos amplificados

Após quantificação dos produtos de RT-PCR, transferir 5µl do produto amplificado para um novo tubo de reacção e adicionar 2µl de reagente PCR Cleanup (exonuclease 1: recombinante, fosfatase alcalina, tampão). Sellar os tubos e efectuar uma breve centrifugação. Colocar os tubos no termociclador de acordo com os tempos da tabela 4.

Tabela 4: Tempos utilizados na purificação

Temperatura (°C)	Duração (min)
37	15:00
80	15:00
4	Manter

Após purificação diluir os produtos de RT-PCR com água destilada e desionizada, antes de iniciar os ciclos de sequenciação. Para amostras amplificadas de acordo com este protocolo, efectuar as diluições de acordo com a tabela 5. Para amostras com uma fraca amplificação (< 20 ng num gel de agarose a 1%) ajustar o volume para 40 µl.

Tabela 5: Valores de diluição para produtos de RT-PCR purificados

Intensidade da banda	Diluição
	1:10
>20 ng	7µl de produto RT-PCR purificado com 47 µl de água pura
	1:4
20 ng	7µl de produto RT-PCR purificado com 33 µl de água pura

### 3.2.7 Reacção de sequenciação dos produtos purificados do gene da integrase.

Colocar os quatro *primers* de sequenciação, HIV IN SEQ MIX A, HIV IN SEQ MIX B, HIV IN SEQ MIX C e HIV IN SEQ MIX D, constituídos por uma mistura de BigDye® Terminator com uma concentração de oligonucleotidos sintéticos não infeccioso inferior a 0,1%, a temperatura ambiente e protegidos da luz (Tabela 6).

Tabela 6:*Primers*

<i>Primer</i>	Posição na região do gene <i>pol</i> (integrase)
A directo	[4484-5119]
B directo	[4215-4857]
C inverso	[4385-5027]
D inverso	[4159-4753]

Usam-se os quatro *primers* para se obter a cobertura máxima da região da integrase do gene *pol*. Adicionar 12 µl de cada *primer* a 8µl do produto purificado e diluído, selar a placa e efectuar uma breve centrifugação. A figura 19 exemplifica a forma como os produtos de PCR e os *primers* são distribuídos numa placa de PCR.

Primer A (12µl) + (8µl) PCR Amostra 1	Primer A (12µl) + (8µl) PCR Amostra 3	Primer A (12µl) + (8µl) PCR Amostra 5
Primer B (12µl) + (8µl) PCR Amostra 1	Primer B (12µl) + (8µl) PCR Amostra 3	Primer B (12µl) + (8µl) PCR Amostra 5
Primer C (12µl) + (8µl) PCR Amostra 1	Primer C (12µl) + (8µl) PCR Amostra 3	Primer C (12µl) + (8µl) PCR Amostra 5
Primer D (12µl) + (8µl) PCR Amostra 1	Primer D (12µl) + (8µl) PCR Amostra 3	Primer D (12µl) + (8µl) PCR Amostra 5
Primer A (12µl) + (8µl) PCR Amostra 2	Primer A (12µl) + (8µl) PCR Amostra 4	Primer A (12µl) + (8µl) PCR Amostra 6
Primer B (12µl) + (8µl) PCR Amostra 2	Primer B (12µl) + (8µl) PCR Amostra 4	Primer B (12µl) + (8µl) PCR Amostra 6
Primer C (12µl) + (8µl) PCR Amostra 2	Primer C (12µl) + (8µl) PCR Amostra 4	Primer C (12µl) + (8µl) PCR Amostra 6
Primer D (12µl) + (8µl) PCR Amostra 2	Primer D (12µl) + (8µl) PCR Amostra 4	Primer D (12µl) + (8µl) PCR Amostra 6

**Figura 19:Exemplo de preenchimento de uma placa para sequenciação**

Efectuar os ciclos de sequenciação de acordo com o protocolo Viroseq™ HIV-1 Integrase Genotyping Kit como consta na tabela 7.

**Tabela 7:Ciclos utilizados na amplificação do gene da integrase**

Temperatura (°C)	Duração	Nº de ciclos
96	00:10	25
50	00:05	
60	4:00	
4	Manter	-

### 3.2.8 Detecção e leitura da sequência de ADN.

Ressuspender os produtos provenientes dos ciclos de sequenciação em 20 µl de formamida HiDi. Selar a placa, efectuar um vortex 10-15 seg., e efectuar uma breve centrifugação. O sequenciador automático de 6 capilares ABI PRISM® 3100 da Applied Biosystems detecta a fluorescência emitida por quatro terminadores marcados com fluorocromos diferentes, usados para identificar as bases de Adenina, Citosina, Guanina, e Timina nas reacções de extensão/terminação ADN. Cada tipo de fluorocromo emite uma luz, num comprimento de onda diferente, quando excitado por um laser de iões de Árgon. Os quatro fluorocromos, ou seja, as quatro bases, são detectados e distinguidos num único capilar. Os produtos purificados são ressuspensos em 20 µl de formamida Hi-Di™. As amostras são então arrastadas electro-cineticamente para os capilares. A sequenciação envolve a criação de uma corrente eléctrica de iões de carga negativa para carga positiva através de um polímero POP-médio 6™. A rede de carga negativa criada durante o ciclo de sequenciação permite que os fragmentos ADN migrem através deste campo, os fragmentos com menor tamanho migram mais rapidamente do que os fragmentos maiores. Cada fragmento passa pela região de leitura do laser, os fluorocromos são excitados pelo ião Árgon do laser. A fluorescência emitida é detectada por uma câmara CCD e convertida numa sequência *basecall* pelo ABI PRISM® *Sequencing Analysis Software*. Os ficheiros resultantes contêm a informação sobre a sequência de cada um dos quatro *primers* para cada produto sequenciado. Após a conclusão da sequenciação os ficheiros obtidos são transferidos para um programa de edição de sequências.



### 3.2.9 Edição das sequências obtidas.

Os ficheiros com as sequências obtidas no ABI PRISM ® 3100 são analisadas e editadas usando o software do SeqScape ® v2.5. A figura 20 exemplifica os resultados obtidos na sequenciação automática. Vemos os electroferogramas dos quatro *primers* utilizados em que cada pico apresenta uma cor diferente (vermelho, azul, verde, preto) correspondente aos diferentes nucleotidos (vermelho=T, azul=C, verde=A, preto=G).

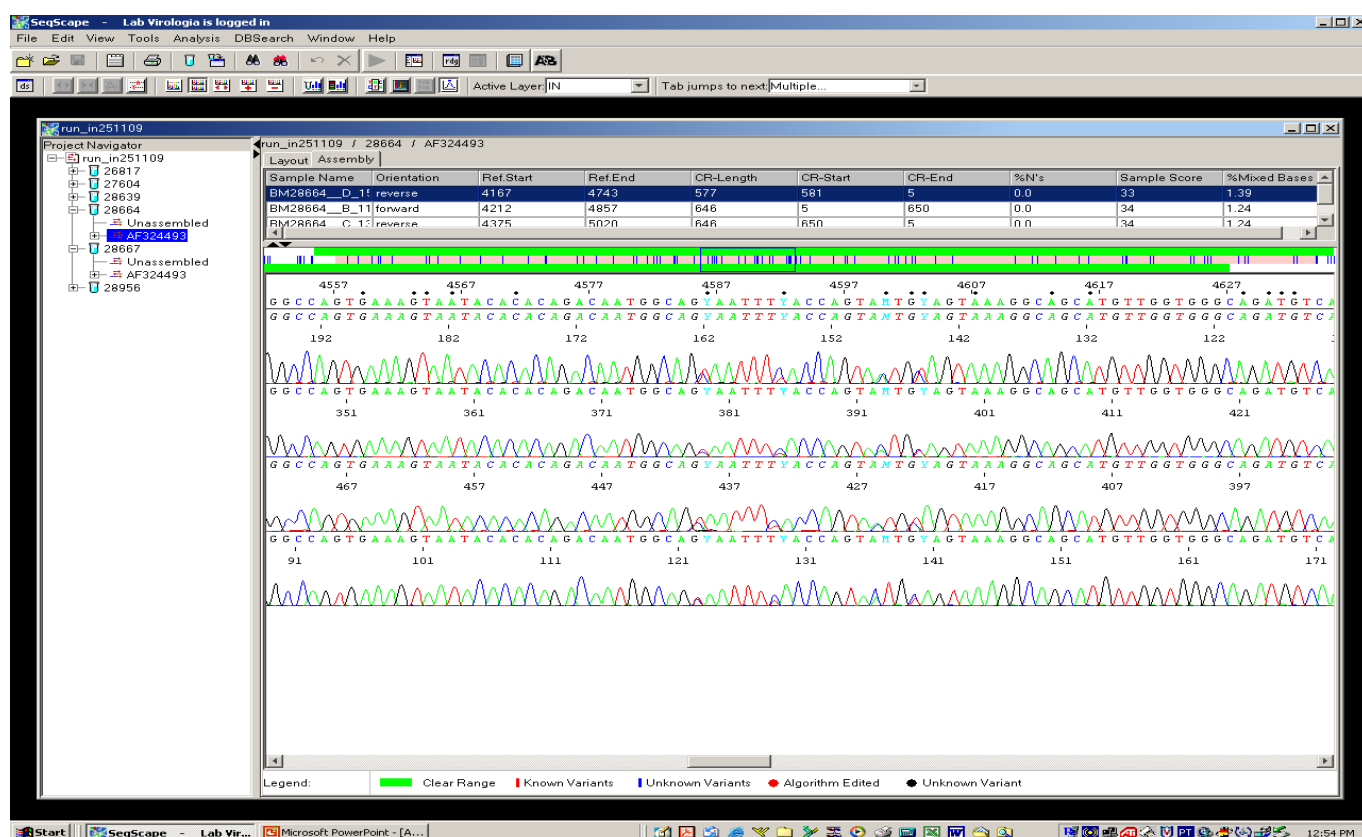
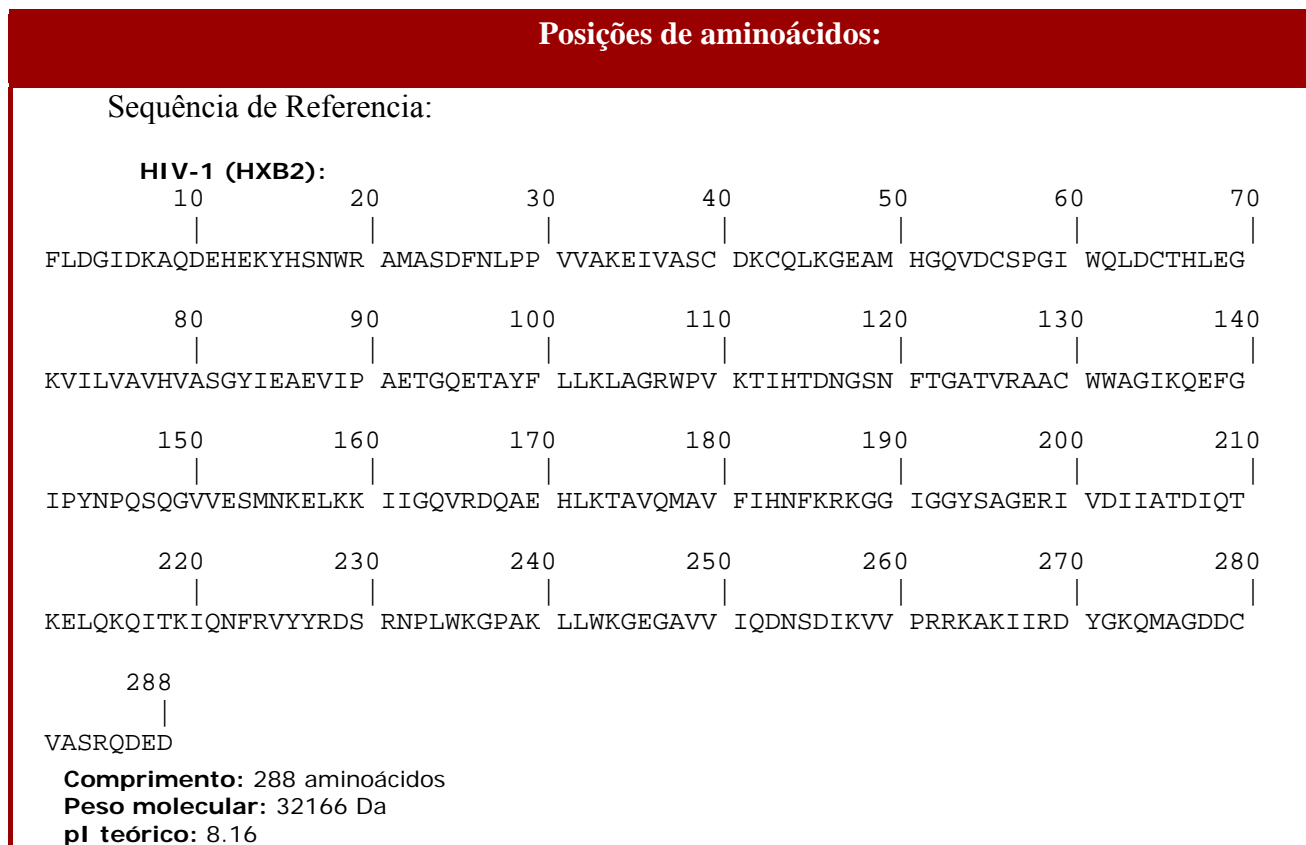


Figura 20: Edição de uma sequência

Após analisar as sequências, o software gera um ficheiro de aminoácidos resultante dessa análise e um ficheiro Fasta com a sequência de nucleótidos. O controlo de qualidade das sequências analisadas é efectuado através da submissão do ficheiro Fasta, à base de dados de Stanford em <http://hivdb.stanford.edu/>. Os aminoácidos alterados são obtidos por comparação

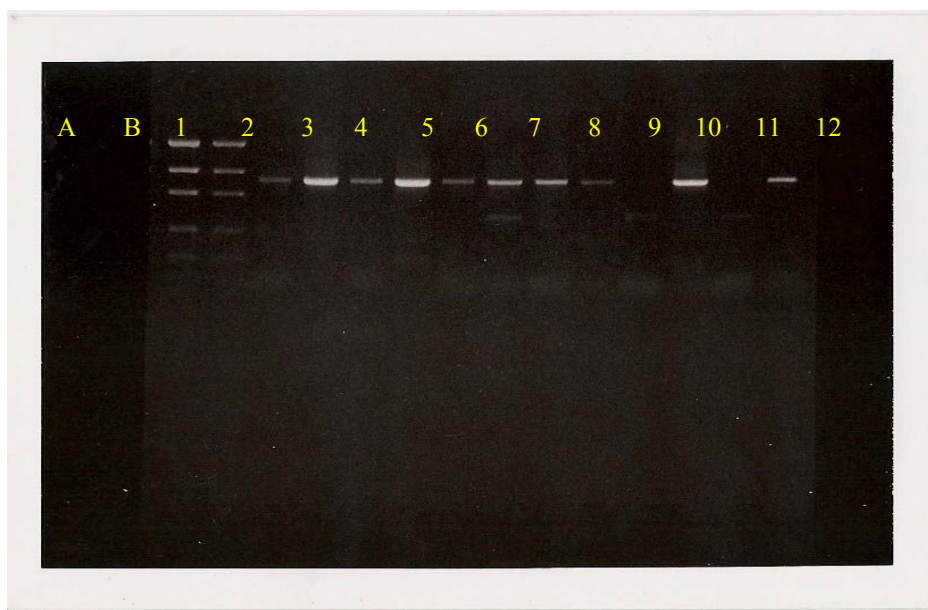
com a sequência de referência HXB2 (figura 21) do GenBank com o número de acesso K03455.



**Figura 21: Alinhamento da sequência de referência HXB2.** Adaptado de: [www.bioafrica.net/bioinformaticstools.html](http://www.bioafrica.net/bioinformaticstools.html)

## 4 Resultados

Foram extraídas 126 amostras, conseguindo-se amplificação em 111 (88%). Partindo dos produtos amplificados obtiveram-se 103 sequências (93%). A taxa de não amplificação foi de 12% e a de insucesso na sequenciação foi de 7%. Isto deve-se provavelmente à qualidade de algumas amostras e à provável presença de inibidores. Os produtos amplificados foram visualizados e quantificados utilizando um gel de agarose a 1%, como se pode observar na figura 22.



**Figura 22: Fotografia das bandas dos produtos amplificados.**

A banda A corresponde ao marcador de peso molecular (6 $\mu$ l), a B ao marcador de peso molecular (3  $\mu$ l), a banda 12 ao controlo positivo, as bandas de 1 a 11 aos produtos amplificados dos doentes VIH-1. As bandas 9 e 11 não apresentam amplificação (negativas). De acordo com a tabela os produtos amplificados apresentam os seguintes pesos: amostra 1=20ng, amostra 2 $\geq$ 100ng, amostra 3=30ng, amostra 4 $\geq$ 100ng, amostra 5=20ng, amostra 6=60ng, amostra 7=60ng, amostra 8=20ng, amostra 9=0ng, amostra 10 $\geq$ 100ng, amostra 11=0ng, amostra 12=60ng.

Após a análise das 103 sequências, verificou-se que das 288 posições da integrase, 96 (33%) posições são polimórficas. Destas posições polimórficas, 21/50 (42%) situam-se na região N-terminal, 52/162 (32%) situam-se na região central e 24/75 (31,5%) situam-se na região C-terminal.

As 103 amostras sequenciadas para a região da integrase do gene *pol* foram agrupadas em seis subtipos obtidos por sequenciação da região da transcriptase reversa e da protease do gene *pol*: subtipo A, 3 sequências; subtipo B, 12 sequências; subtipo C, 12 sequências; subtipo F, 5 sequências; subtipo G, 48 sequências; subtipo CRFO2-AG, 23 sequências.

#### **4.1 Análise da região N-terminal.**

Para além do arranjo conservado HHCC ( $H_{12}H_{16}C_{40}C_{43}$ ) existem outros aminoácidos conservados de forma isolada ou em pequenos trechos (F1-L2,G4-I5,A8-Q9,Y15-H16,N18-W19,E35I36,C43-Q44,K46-A49). Salienta-se que a posição 12, que tem sido alvo de experiências *in vivo*, neste estudo apresentou-se conservada.

#### **4.2 Análise da região central.**

Na região mais conservada da integrase, a maioria dos aminoácidos conservados (112/163, 68%) distribuem-se de uma forma isolada ou agrupados em quatro domínios distintos de grande conservação. Estes domínios são constituídos por 7 a 14 aminoácidos (figura 23) e contêm alguns aminoácidos funcionais importantes:

- I (W61-K71),
- II (V75-Y83),
- III (K127-A133),
- IV (Q137-V150).

	F1	L2	D3	G4	I5	D6	K7	A8	Q9	E10	E11	H12	E13	K14	Y15
	H16	S17	N18	W19	R20	A21	M22	A23	S24	D25	F26	N27	L28	P29	P30
	V31	V32	A33	K34	E35	I36	V37	A38	S39	C40	D41	K42	C43	Q44	L45
	K46	G47	E48	A49	M50	H51	G52	Q53	V54	D55	C56	S57	P58	G59	I60
I	W61	Q62	L63	D64	C65	T66	H67	L68	E69	G70	K71	I72	I73	L74	V75
II	A76	V77	H78	V79	A80	S81	G82	Y83	I84	E85	A86	E87	V88	I89	P90
	A91	E92	T93	G94	Q95	E96	T97	A98	Y99	F100	L101	L102	K103	L104	A105
	G106	R107	W108	P109	V110	K111	T112	H113	H114	T115	D116	N117	G118	S119	N120
III	F121	T122	S123	T124	T125	V126	K127	A128	A129	C130	W131	W132	A133	G134	I135
IV	K136	Q137	E138	F139	G140	I141	P142	Y143	N144	P145	Q146	S147	Q148	G149	V150
	V151	E152	S153	M154	N155	K156	E157	L158	K159	K160	I161	I162	G163	Q164	V165
	R166	D167	Q168	A169	E170	H171	L172	K173	T174	A175	V176	Q177	M178	A179	V180
	F181	I182	H183	N184	F185	K186	R187	K188	G189	G190	I191	G192	G193	Y194	S195
	A196	G197	E198	R199	I200	V201	D202	I203	I204	A205	T206	I207	I208	Q209	T210
	K211	E212	L213	Q214	K215	Q216	I217	T218	K219	I220	Q221	N222	F223	R224	V225
	Y226	Y227	R228	D229	S230	R231	D232	P233	L234	W235	K236	G237	P238	A239	K240
V	L241	L242	W243	K244	G245	E246	G247	A248	V249	V250	I251	Q252	D253	N254	S255
VI	D256	I257	K258	V259	V260	P261	R262	R263	K264	A265	K266	I267	I268	R269	D270
	Y271	G272	K273	Q274	M275	A276	G277	D278	D279	C280	V281	A282	S283	R284	Q285
	D286	E287	D288												

<1%

1,1-5%

5,1-10%

10,1-25%

>25%

**Figura 23: Resultados percentuais da variabilidade das posições dos aminoácidos**

O domínio I contém o primeiro aminoácido (D64) da tríade catalítica (D<sub>64</sub>D<sub>116</sub>E<sub>152</sub>). O aminoácido altamente conservado D116 encontra-se no pequeno domínio conservado (H114-G118). Neste estudo, para além da conservação do centro catalítico, também os aminoácidos P109 e T115 estão altamente conservados. No domínio III, o aminoácido C130 é essencial à manutenção da estrutura terciária da integrase, aqui também se encontra altamente conservado.

O maior domínio conservado (IV,137-150) contém o dito laço “catalítico” G140-G149, que se encontra conservado. Outros aminoácidos deste grande domínio tais como, E138, G140, Y143, Q146, S147, Q148 associados à resistência aos inibidores da integrase, encontram-se também conservados.

O arranjo KRK (K<sub>186</sub> R<sub>187</sub> K<sub>188</sub>) importante para oligomerização na interface dímero: dímero da integrase não está totalmente conservado. Só o aminoácido K186 está altamente

conservado. As posições 187 e 188 têm uma variabilidade entre 1,1 a 2,9%. Na região central existe uma sequência de aminoácidos altamente conservada (I161-K173) (Bouyac-Bertoia et al. 2001a). Esta sequência de aminoácidos funciona como sinalizador de localização nuclear (SNL). Neste estudo, apesar de estar altamente conservada existem duas exceções as posições 163 e 165 com uma variabilidade 1 e 7% respectivamente. Estas duas posições estão associadas à resistência a diferentes inibidores, tanto *in vitro* como *in vivo*.

### 4.3 Análise da região C-terminal

Esta região está envolvida na ligação ao ADN celular e viral e contém dois domínios, o V e o VI. O domínio V é constituído pelos aminoácidos conservados (L241-Q252) e o domínio VI pelos aminoácidos (I257-K264)). Os dois domínios estão quase sobrepostos à área C (<sup>235</sup>WKGPAKLLWGE<sup>250</sup>) e à área N (<sup>259</sup>VVPRRK<sup>264</sup>), que são áreas conservadas em todos os retrovírus e essenciais à replicação viral (Cannon et al. 1996c). A área C (<sup>235</sup>W-V<sup>250</sup>) está altamente conservada com excepção da posição 249 (1% de variabilidade). Outras mutações como a E246K e G247W que se sabe estarem associadas a uma alteração do “splicing “ do ARN, ao processamento das proteínas GAG e à produção de partículas virais estão ausentes (Cannon *et al.* 1996b).

No domínio conservado VI (I257-K264) existem vários aminoácidos importantes, tal como o V260 importante para multimerização da integrase (Kalpana *et al.* 1999). Os aminoácidos K258,R262,R263 E K264 estão relacionados com a ligação ao ADN. (Dirac and Kjems 2001). Células infectadas com VIH-1 com a mutação V260E na integrase não conseguem ter níveis de replicação viral detectável. Os aminoácidos E246 e K273 também relacionados com a ligação ao ADN estão altamente conservados. Neste estudo, a posição 260 apresenta uma variabilidade de 1%.

## 4.4 Análise geral

A figura 23 representa a distribuição das 288 posições da integrase. As 14 posições assinaladas a vermelho, correspondem a posições com taxas de variabilidade  $\geq$  a 25%. As posições 151 e 201 estão associadas a resistência em doentes medicados com Raltegravir, (Rhee et al. 2008) neste estudo apresentam-se como polimorfismos naturais. Todas as posições associadas a resistência aos inibidores da integrase estão sublinhadas e a sua variabilidade, situa-se entre 1 e/ou superior a 25%. Os aminoácidos L74 e T97 associados a resistência secundária apresentam uma variabilidade entre (1 e 1,9%) e o aminoácido S119 (10,1 a 25%). Estas posições são seleccionadas nos doentes sob terapêutica com Raltegravir, (Rhee et al. 2008) neste estudo apresentam-se como polimorfismos naturais. Neste estudo, foram identificadas taxas de polimorfismos superiores a 25% em 5,3% das posições da região N-terminal, 4,3% na região Central e 32% na região C-terminal (esta é a região menos conservada da integrase). As mutações primárias, já descritas na literatura, associadas a um alto grau de resistências (N155H,Y143R,Q148H/K/R) (Rhee et al. 2008) não foram encontradas neste estudo.

## 5 Co-variação do subtipo G do VIH-1

### 5.1 Importância do subtipo G em Portugal

Os subtipos A, B e C são os subtipos com maior prevalência a nível mundial de acordo com estudos recentes (Buonoaguro *et al*):

- Subtipo B, 10,42%
- Subtipo A, 12,3%
- Subtipo C, 49,9%

- Subtipo G, 6,32%
- Subtipo F, 0,59%
- Subtipo D, 2,53%
- Subtipo K, 0,04%
- Subtipo JO, 0,14%
- Subtipo CRFs, 17,58%

Em Portugal, a prevalência dos subtipos (B e G) num estudo realizado por Palma e colaboradores com doentes infectados pelo VIH-1, identificaram o subtipo B com uma prevalência de (41,7%) e o subtipo G como o segundo subtipo com maior prevalência (29,4%) (Palma et al. 2007).

No estudo aqui apresentado o subtipo G foi o mais prevalente, por esta razão e por existirem poucos estudos sobre este subtipo, foi-lhe dada especial atenção no intuito de contribuir para um melhor conhecimento do subtipo G.

Na amostra estudada foram identificadas as seguintes variabilidades entre os subtipos analisados:

- Subtipo G, 51 posições polimórficas (17,7%),
- Subtipo B, 54 posições polimórficas ( 17,3%),
- Subtipo CRFO-2AG, 48 posições polimórficas (16,6%)
- Subtipo C, 47 posições polimórficas (16,3%),
- Subtipo F, 38 posições polimórficas (13,2%),



- Subtipo A, 19 posições polimórficas (6,5%).

O subtipo G foi o grupo com maior número de sequências analisadas (48), correspondendo ao que apresentou o maior número de polimorfismos (17,7%) relativamente à sequência de referência (HXB<sub>2</sub>). A análise estatística da co-variação entre aminoácidos foi efectuada através do programa SPSS (*Statistical Package for the Social Sciences*) versão 16.0. Para análise de possíveis associações entre pares de aminoácidos polimórficos utilizou-se o índice de similaridade de Jaccard (J) (Turechek 2004). Os valores do índice de Jaccard variam entre 0 e 1. Os valores próximos de 1 indicam um alto grau de similaridade, enquanto os valores próximos de 0 indicam uma dissociação. O índice de Jaccard é calculado utilizando a fórmula:  $J = N_{xy} / (N_{xy} + N_{xo} + N_{oy})$ , onde  $N_{xy}$  representa o número de sequências com ambos os polimorfismos,  $N_{xo}$  representa o número de sequências com a presença de X e ausência de Y,  $N_{oy}$  representa o número de sequências com a presença de Y e ausência de X. Para testar a significância dos índices de similaridade obtidos, foi calculado o índice de similaridade de Jaccard *random* ( $J_{rand}$ ) e o desvio padrão ( $J_{SE}$ ) assumindo que as duas mutações (X e Y) ocorrem independentemente. O  $J_{rand}$  foi calculado como sendo a média do índice de similaridade de Jaccard após 1000 rearranjos entre X e Y. O  $J_{SE}$  foi calculado utilizando o método de Jack-knifed (Turechek and Madden 2000). Desta forma, foi possível padronizar os valores de Z através da fórmula  $Z = (J - J_{rand}) / J_{SE}$ , os valores de Z foram tratados como um teste bi-caudal em que os valores de  $Z > 1,96$  indicam uma associação positiva estatisticamente significativa, e valores de  $Z < -1,96$  indicam uma associação negativa ou dissociação estatisticamente significativa para um valor  $\alpha = 0,05$ .

**Tabela 8 :Valores de associação índice de Jaccard**

<b>Pares de aminoácidos polimórficos</b>	<b>Valor J</b>	<b>Valor Z</b>	<b>Valor P</b>
<b>D11/I31</b>	0,88	2,42	0,0
<b>R14/I112</b>	0,97	9,66	0,0
<b>R14/N134</b>	0,89	4,16	0,0
<b>R14/E256</b>	0,84	2,90	0,0
<b>R14/G283</b>	0,86	3,45	0,0
<b>I72/N134</b>	0,80	2,18	0,0
<b>I72/V151</b>	0,94	9,28	0,0
<b>I72/I201</b>	0,98	14,33	0,0
<b>I72/I234</b>	0,96	9,28	0,0
<b>I72/N255</b>	0,91	5,68	0,0
<b>I72/E256</b>	0,84	3,22	0,0
<b>I72/G283</b>	0,82	2,67	0,0
<b>I101/N134</b>	0,80	2,18	0,0
<b>I101/V151</b>	0,94	9,28	0,0
<b>I101/I201</b>	0,98	14,33	0,0
<b>I101/I234</b>	0,96	9,28	0,0
<b>I101/N255</b>	0,91	5,68	0,0
<b>I101/E256</b>	0,84	3,22	0,0
<b>I101/G283</b>	0,82	2,67	0,0
<b>I112/N134</b>	0,86	3,30	0,0
<b>I112/F227</b>	0,97	9,66	0,0
<b>I112/E256</b>	0,82	2,31	0,0
<b>I112/G283</b>	0,84	2,76	0,0
<b>A124/I72</b>	0,88	4,87	0,0
<b>A124/I101</b>	0,88	4,87	0,0
<b>A124/A125</b>	0,98	15,6	0,0
<b>A124/T136</b>	0,88	4,87	0,0
<b>A124/V151</b>	0,94	8,26	0,0
<b>A124/I201</b>	0,90	5,99	0,0
<b>A124/S206</b>	0,88	4,87	0,0
<b>A124/I234</b>	0,83	3,32	0,0
<b>A124/N255</b>	0,79	2,27	0,0
<b>A125/I72</b>	0,85	4,87	0,0
<b>A125/I101</b>	0,85	4,87	0,0
<b>A125/T136</b>	0,85	4,87	0,0
<b>A125/V151</b>	0,92	8,26	0,0
<b>A125/I201</b>	0,88	5,99	0,0
<b>A125/S206</b>	0,85	4,87	0,0
<b>A125/I234</b>	0,81	3,32	0,0
<b>A125/N255</b>	0,77	2,27	0,0
<b>N134/T136</b>	0,80	2,18	0,0
<b>N134/S206</b>	0,80	2,18	0,0
<b>N134/F227</b>	0,89	4,16	0,0
<b>N134/I234</b>	0,84	2,96	0,0
<b>N134/N255</b>	0,88	4,06	0,0
<b>N134/E256</b>	0,95	7,63	0,0

<b>N134/G283</b>	0,97	11,83	0,0
<b>T136/V151</b>	0,94	9,28	0,0
<b>T136/I201</b>	0,98	14,33	0,0
<b>T136/I234</b>	0,96	9,28	0,0
<b>T136/N255</b>	0,91	5,68	0,0
<b>T136/E256</b>	0,84	3,22	0,0
<b>T136/G283</b>	0,82	2,67	0,0
<b>V151/I201</b>	0,96	13,99	0,0
<b>V151/S206</b>	0,94	9,28	0,0
<b>V151/I234</b>	0,90	5,66	0,0
<b>V151/N255</b>	0,85	3,86	0,0
<b>V151/E256</b>	0,79	2,16	0,0
<b>I201/S206</b>	0,98	14,33	0,0
<b>I201/I234</b>	0,93	7,27	0,0
<b>I201/N255</b>	0,89	4,82	0,0
<b>I201/E256</b>	0,83	2,79	0,0
<b>I201/G283</b>	0,80	2,29	0,0
<b>S206/I234</b>	0,96	9,28	0,0
<b>S206/N255</b>	0,91	5,68	0,0
<b>S206/E256</b>	0,84	3,22	0,0
<b>S206/G283</b>	0,82	2,67	0,0
<b>F227/E256</b>	0,84	2,90	0,0
<b>F227/G283</b>	0,86	3,45	0,0
<b>I234/N255</b>	0,95	8,81	0,0
<b>I234/E256</b>	0,88	4,36	0,0
<b>I234/G283</b>	0,86	3,60	0,0
<b>N255/E256</b>	0,93	6,30	0,0
<b>N255/G283</b>	0,90	5,00	0,0
<b>E256/G283</b>	0,97	11,00	0,0

Estes pares de aminoácidos apresentam uma forte associação (0,79-0,97) com um valor  $P=0,0$ . Estas posições são descritas por Rhee e colaboradores como sendo as posições da sequência consensus do subtipo G (Rhee et al. 2008). As posições D11, I72, V151 não foram identificadas nas 35 sequências do subtipo G analisadas por Rhee e colaboradores. No estudo efectuado por Rhee a posição 112 está associada a uma valina, no caso do nosso subtipo G (português) essa mesma posição está associada a uma isoleucina. Estes quatro aminoácidos poderão constituir um perfil característico do subtipo G existente em Portugal.

## 6 Discussão

Muitos aspectos da integração e da inibição da integrase permanecem por esclarecer. O posicionamento relativo, entre as três regiões e a estrutura tri-dimensional e multimérica da forma activa da integrase, continua em estudo.

Nos últimos 10 anos, o número de artigos que descrevem a forma como as mutações são seleccionadas pelos inibidores, *in vitro* e *in vivo* e a forma como essas mutações afectam a sensibilidade aos inibidores, forneceram muitos conceitos sobre a resistência aos inibidores da integrase:

- 1º. Muitas dessas mutações foram seleccionadas *in vitro* ou *in vivo*,
- 2º. As mutações que reduzem a susceptibilidade aos inibidores da integrase ocorrem na proximidade da tríade catalítica,
- 3º. Muitas das mutações são secundárias uma vez que pouco ou nenhum efeito tem sobre a susceptibilidade aos inibidores,
- 4º. A falência terapêutica com o Raltegravir ou com Elvitegravir foi acompanhada por uma ou duas mutações primárias e uma perda na ordem dos 10 a 100 vezes na susceptibilidade aos inibidores.

As mutações primárias (Y143R, Q148H/K/R, N155H), que conferem resistência de alto grau ao Raltegravir estão ausentes neste estudo tal como no estudo de (Rhee *et al.* 2008) que analisou mais de 1500 sequencias do grupo M provenientes de indivíduos que nunca tomaram inibidores da integrase, reportando que 115 das 288 posições (39.9%) da integrase são polimórficas e que 202 das 288 posições (70%) estão conservadas. As mutações secundárias (H51Y, T66A/K, E138K, G140S/A/C, K160N, R166S, E170A, S230R e R263K) estão ausentes

neste estudo. Algumas mutações secundárias identificadas em doentes com falência terapêutica ao Raltegravir, tais como a (L68I, E138A, E157Q, G163K/R, 232N) estão ausentes. No entanto, a posição 163 apresenta uma variabilidade de 6,7%, valor semelhante ao encontrado no estudo efectuado por Ceccherini-Silberstein e colaboradores (Ceccherini-Silberstein et al. 2009b) que analisou 448 subtipos B provenientes respectivamente de 134 indivíduos que nunca tomaram inibidores da integrase e de 314 indivíduos em falência terapêutica mas sem terapêutica aos inibidores da integrase.

O arranjo HHCC, que estabelece a ligação ao zinco, promove a multimerização das sub-unidades da integrase e a estabilização da enzima, está conservado. (Burke *et al.* 1992). O HHCC também está envolvido na interacção com o co-factor LEDGF/p75. A mutação do H12N implica uma perda da afinidade para o co-factor acima mencionado. Alterações tais como H12N, H12A, H16A e C40S originam vírus defectivos do ponto de vista da replicação e uma grande diminuição do processamento 3' e transferência de cadeias (Engelman *et al.* 1995). Tais mutações não foram encontradas neste estudo. De forma semelhante a conservação dos aminoácidos A33, E35-I36 são importantes para actividade catalítica (processamento 3' e transferência de cadeias). Lu e colaboradores demonstraram que mutantes K34A são incapazes de se replicarem, a importância da conservação deste aminoácido é reforçada pelo facto de nunca se ter encontrado *in vivo* (Lu *et al.* 2005). Neste estudo, estas mutações também não foram encontradas. Drelich e colaboradores propõem que a mutação dos aminoácidos conservados D64, P109, D116 e E152 afectam de forma adversa a actividade funcional *in vitro*. A mutação do aminoácido T115 por uma alanina causa uma quase total perda da actividade de integração dependente do  $Mg^{2+}$  (Drelich *et al.* 1993). Um estudo de 1996 demonstrou que integrases com a mutação K186T e K188T, são cataliticamente inactivas em linhas celulares C8166 e H9 (Cannon et al. 1996a). A carga positiva da região é fundamental para estabilidade da estrutura da integrase e da capacidade infecciosa do vírus. No estudo aqui apresentado a

mutação K188T não foi encontrada mas sim a mutação (K188R), com uma variabilidade de 2,9% valores sobreponíveis aos publicados no trabalho de Rhee e colaboradores para a mutação K188T, nos dois tipos de mutação, um aminoácido hidrofílico (K) é substituído por um outro aminoácido hidrofílico (T;R). As duas mutações V165A e R166A que impedem a importação nuclear, formação do provírus e consequentemente a replicação viral (Wiskerchen e Muesing 1995) estão ausentes nos subtipos analisados neste estudo. Células infectadas com VIH-1 com a mutação V260E na integrase não conseguem ter níveis de replicação viral detectável. Os aminoácidos E246 e K273 que participam na ligação ao ADN estão totalmente conservados.

Neste estudo, verificou-se que as mutações primárias associadas a resistência e com significado clínico estão ausentes nos doentes, que não são tratados com inibidores da integrase. Sabe-se que existem diferentes perfis de resistência para a integrase (Y143R, Q148H/K/R, N155H) tal como acontece para as outras classes de anti-retrovirais (IP, NNRTI, NRTI). Contudo, não se sabe o que determina o aparecimento de um perfil de resistência em detrimento de outro e a forma como evoluem os diferentes perfis de resistência da integrase. É possível que a existência prévia de polimorfismos naturais com grande variabilidade possa condicionar a selecção dos perfis de resistência.

Na ausência da pressão de inibidores, a integrase é uma enzima que requer *in vivo* a conservação de 2/3 dos seus aminoácidos. Será necessário no futuro identificar a conservação mínima de aminoácidos necessária na presença de inibidores. A caracterização dos aminoácidos altamente conservados (envolvidos na estabilidade da enzima, multimerização, ligação ao ADN, actividade catalítica, ligação ao LEDGF) poderá ajudar no desenvolvimento de novos inibidores com mecanismos acção alternativos e perfis de resistência mais favoráveis.

O presente estudo, demonstrou a importância da caracterização das posições polimórficas do gene *Pol* da integrase. Este estudo contribuiu para identificação das posições polimórficas em doentes não tratados com inibidores da integrase e confirmar a ausência de mutações associadas a resistência nestes doentes. Este estudo foi importante para verificar que a percentagem de polimorfismos varia de acordo com a região analisada, sendo a região C-terminal a mais variável. A análise mais detalhada do subtipo G contribui de forma significativa para caracterizar os polimorfismos existentes na população de doentes VIH-1 portuguesa e revelar algumas diferenças significativas em relação ao estudo efectuado por Rhee e colaboradores com o mesmo subtipo G. A variabilidade dos polimorfismo intra-subtipos e inter-subtipos dificultam a determinação dos possíveis perfis de resistência. Este estudo contribui ainda para a criação de uma base de dados que irá permitir no futuro avaliar a evolução dos polimorfismos nos diferentes subtipos. A utilização do índice de Jaccard permitiu de forma inequívoca demonstrar a associação entre posições polimórficas e verificar como essas associações poderão evoluir. A partir deste estudo estão criadas as condições para um melhor conhecimento do gene *pol* da integrase nos diferentes subtipos.

## 7 Bibliografia

1. **Andrake, M. D. and A. M. Skalka. 1995.** Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. *J. Biol. Chem.* 270:29299-29306.
2. **Ao, Z., G. Huang, H. Yao, Z. Xu, M. Labine, A. W. Cochrane, and X. Yao. 2007.** Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J. Biol. Chem.* 282:13456-13467.
3. **Asante-Appiah, E., S. H. Seeholzer, and A. M. Skalka. 1998.** Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J. Biol. Chem.* 273:35078-35087.
4. **Berthoux, L., S. Sebastian, M. A. Muesing, and J. Luban. 2007.** The role of lysine 186 in HIV-1 integrase multimerization. *Virology* 364:227-236.
5. **Blankson, J. N., D. Persaud, and R. F. Siliciano. 2002.** The challenge of viral reservoirs in HIV-1 infection. *Annu. Rev. Med.* 53:557-593.
6. **Bouyac-Bertoia, M., J. D. Dvorin, R. A. Fouchier, Y. Jenkins, B. E. Meyer, L. I. Wu, M. Emerman, and M. H. Malim. 2001a.** HIV-1 infection requires a functional integrase NLS. *Mol. Cell* 7:1025-1035.
7. **Bouyac-Bertoia, M., J. D. Dvorin, R. A. Fouchier, Y. Jenkins, B. E. Meyer, L. I. Wu, M. Emerman, and M. H. Malim. 2001b.** HIV-1 infection requires a functional integrase NLS. *Mol. Cell* 7:1025-1035.
8. **Bujacz, G., M. Jaskolski, J. Alexandratos, A. Wlodawer, G. Merkel, R. A. Katz, and A. M. Skalka. 1996.** The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure.* 4:89-96.
9. **Bukrinsky, M. I., N. Sharova, M. P. Dempsey, T. L. Stanwick, A. G. Bukrinskaya, S. Haggerty, and M. Stevenson. 1992.** Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. U. S. A* 89:6580-6584.
10. **Buonaguro, L., M. L. Tornesello, and F. M. Buonaguro. 2007.** Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* 81:10209-10219.
11. **Burke, C. J., G. Sanyal, M. W. Bruner, J. A. Ryan, R. L. LaFemina, H. L. Robbins, A. S. Zeff, C. R. Middaugh, and M. G. Cordingley. 1992.** Structural



- implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.* 267:9639-9644.
12. **Busschots, K., A. Voet, M. De Maeyer, J. C. Rain, S. Emiliani, R. Benarous, L. Desender, Z. Debyser, and F. Christ. 2007.** Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J. Mol. Biol.* 365:1480-1492.
  13. **Cai, M., R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn. 1997.** Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4:567-577.
  14. **Cannon, P. M., E. D. Byles, S. M. Kingsman, and A. J. Kingsman. 1996a.** Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J. Virol.* 70:651-657.
  15. **Cannon, P. M., E. D. Byles, S. M. Kingsman, and A. J. Kingsman. 1996b.** Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J. Virol.* 70:651-657.
  16. **Cannon, P. M., E. D. Byles, S. M. Kingsman, and A. J. Kingsman. 1996c.** Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J. Virol.* 70:651-657.
  17. **Carlson, H. A., K. M. Masukawa, K. Rubins, F. D. Bushman, W. L. Jorgensen, R. D. Lins, J. M. Briggs, and J. A. McCammon. 2000.** Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43:2100-2114.
  18. **Ceccherini-Silberstein, F., I. Malet, R. D'Arrigo, A. Antinori, A. G. Marcelin, and C. F. Perno. 2009a.** Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev.* 11:17-29.
  19. **Ceccherini-Silberstein, F., I. Malet, R. D'Arrigo, A. Antinori, A. G. Marcelin, and C. F. Perno. 2009b.** Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev.* 11:17-29.
  20. **Cherepanov, P., G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, and Z. Debyser. 2003.** HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278:372-381.
  21. **Coffin, J. M. 1995.** HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267:483-489.
  22. **Colson, P., M. Henry, N. Tivoli, H. Gallais, J. A. Gastaut, J. Moreau, and C. Tamalet. 2005.** Polymorphism and drug-selected mutations in the reverse transcriptase gene of HIV-2 from patients living in southeastern France. *J. Med. Virol.* 75:381-390.

23. **De Clercq, E. 2002.** New anti-HIV agents and targets. *Med. Res. Rev.* 22:531-565.
24. **Deeks, S. G. 2001.** International perspectives on antiretroviral resistance. Nonnucleoside reverse transcriptase inhibitor resistance. *J. Acquir. Immune. Defic. Syndr.* 26 Suppl 1:S25-S33.
25. **Dirac, A. M. and J. Kjems. 2001.** Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur. J. Biochem.* 268:743-751.
26. **Drelich, M., M. Haenggi, and J. Mous. 1993.** Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate. *J. Virol.* 67:5041-5044.
27. **Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davies. 1994.** Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266:1981-1986.
28. **Engelman, A. and R. Craigie. 1992.** Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* 66:6361-6369.
29. **Engelman, A. and R. Craigie. 1995.** Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J. Virol.* 69:5908-5911.
30. **Engelman, A., G. Englund, J. M. Orenstein, M. A. Martin, and R. Craigie. 1995.** Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J. Virol.* 69:2729-2736.
31. **Espeseth, A. S., P. Felock, A. Wolfe, M. Witmer, J. Grobler, N. Anthony, M. Egbertson, J. Y. Melamed, S. Young, T. Hamill, J. L. Cole, and D. J. Hazuda. 2000.** HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl. Acad. Sci. U. S. A* 97:11244-11249.
32. **Esposito, D. and R. Craigie. 1998.** Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* 17:5832-5843.
33. **Farnet, C. M. and F. D. Bushman. 1997.** HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* 88:483-492.
34. **Fikkert, V., B. Van Maele, J. Vercammen, A. Hantson, B. Van Remoortel, M. Michiels, C. Gurnari, C. Pannecouque, M. De Maeyer, Y. Engelborghs, E. De Clercq, Z. Debyser, and M. Witvrouw. 2003a.** Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations. *J. Virol.* 77:11459-11470.

35. **Fikkert, V., B. Van Maele, J. Vercammen, A. Hantson, B. Van Remoortel, M. Michiels, C. Gurnari, C. Pannecouque, M. De Maeyer, Y. Engelborghs, E. De Clercq, Z. Debyser, and M. Witvrouw. 2003b.** Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations. *J. Virol.* 77:11459-11470.
36. **Fisher, A. G., B. Ensoli, D. Looney, A. Rose, R. C. Gallo, M. S. Saag, G. M. Shaw, B. H. Hahn, and F. Wong-Staal. 1988.** Biologically diverse molecular variants within a single HIV-1 isolate. *Nature* 334:444-447.
37. **Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, and A. Saxon. 1981.** Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N. Engl. J. Med.* 305:1425-1431.
38. **Grindley, N. D. and A. E. Leschziner. 1995.** DNA transposition: from a black box to a color monitor. *Cell* 83:1063-1066.
39. **Grobler, J. A., K. Stillmock, B. Hu, M. Witmer, P. Felock, A. S. Espeseth, A. Wolfe, M. Egbertson, M. Bourgeois, J. Melamed, J. S. Wai, S. Young, J. Vacca, and D. J. Hazuda. 2002.** Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci. U. S. A* 99:6661-6666.
40. **Hammer, S. M., J. J. Eron, Jr., P. Reiss, R. T. Schooley, M. A. Thompson, S. Walmsley, P. Cahn, M. A. Fischl, J. M. Gatell, M. S. Hirsch, D. M. Jacobsen, J. S. Montaner, D. D. Richman, P. G. Yeni, and P. A. Volberding. 2008.** Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. *JAMA* 300:555-570.
41. **Harper, A. L., M. Sudol, and M. Katzman. 2003.** An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. *J. Virol.* 77:3838-3845.
42. **Hazuda, D. J., P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, and M. D. Miller. 2000.** Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* 287:646-650.
43. **Hindmarsh, P., T. Ridky, R. Reeves, M. Andrade, A. M. Skalka, and J. Leis. 1999.** HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. *J. Virol.* 73:2994-3003.
44. **Hu, W. S. and H. M. Temin. 1990a.** Retroviral recombination and reverse transcription. *Science* 250:1227-1233.

45. **Hu, W. S. and H. M. Temin. 1990b.** Retroviral recombination and reverse transcription. *Science* 250:1227-1233.
46. **Jacque, J. M. and M. Stevenson. 2006.** The inner-nuclear-envelope protein emerlin regulates HIV-1 infectivity. *Nature* 441:641-645.
47. **Jegede, O., J. Babu, R. Di Santo, D. J. McColl, J. Weber, and M. Quinones-Mateu. 2008.** HIV type 1 integrase inhibitors: from basic research to clinical implications. *AIDS Rev.* 10:172-189.
48. **Jenkins, T. M., A. Engelman, R. Ghirlando, and R. Craigie. 1996.** A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.* 271:7712-7718.
49. **Johnson, A. A., C. Marchand, and Y. Pommier. 2004.** HIV-1 integrase inhibitors: a decade of research and two drugs in clinical trial. *Curr. Top. Med. Chem.* 4:1059-1077.
50. **Johnson, A. A., W. Santos, G. C. Pais, C. Marchand, R. Amin, T. R. Burke, Jr., G. Verdine, and Y. Pommier. 2006.** Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J. Biol. Chem.* 281:461-467.
51. **Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff. 1994b.** Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 266:2002-2006.
52. **Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff. 1994a.** Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 266:2002-2006.
53. **Kalpana, G. V., A. Reicin, G. S. Cheng, M. Sorin, S. Paik, and S. P. Goff. 1999.** Isolation and characterization of an oligomerization-negative mutant of HIV-1 integrase. *Virology* 259:274-285.
54. **Kellam, P. and B. A. Larder. 1995.** Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J. Virol.* 69:669-674.
55. **Khan, E., J. P. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka. 1991.** Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 19:851-860.
56. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. Mack, and A. M. Skalka. 1992.** Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell Biol.* 12:2331-2338.

57. **Lee, S. P. and M. K. Han. 1996.** Zinc stimulates Mg<sup>2+</sup>-dependent 3'-processing activity of human immunodeficiency virus type 1 integrase in vitro. *Biochemistry* 35:3837-3844.
58. **Lewinski, M. K. and F. D. Bushman. 2005.** Retroviral DNA integration--mechanism and consequences. *Adv. Genet.* 55:147-181.
59. **Lins, R. D., A. Adesokan, T. A. Soares, and J. M. Briggs. 2000.** Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacol. Ther.* 85:123-131.
60. **Loveday, C. 2001.** International perspectives on antiretroviral resistance. Nucleoside reverse transcriptase inhibitor resistance. *J. Acquir. Immune. Defic. Syndr.* 26 Suppl 1:S10-S24.
61. **Lu, R., N. Vandegraaff, P. Cherepanov, and A. Engelman. 2005.** Lys-34, dispensable for integrase catalysis, is required for preintegration complex function and human immunodeficiency virus type 1 replication. *J. Virol.* 79:12584-12591.
62. **Mansky, L. M. and H. M. Temin. 1995.** Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* 69:5087-5094.
63. **Marchand, C., X. Zhang, G. C. Pais, K. Cowansage, N. Neamati, T. R. Burke, Jr., and Y. Pommier. 2002b.** Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J. Biol. Chem.* 277:12596-12603.
64. **Marchand, C., X. Zhang, G. C. Pais, K. Cowansage, N. Neamati, T. R. Burke, Jr., and Y. Pommier. 2002a.** Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J. Biol. Chem.* 277:12596-12603.
65. **McDonald, D., M. A. Vodicka, G. Lucero, T. M. Svitkina, G. G. Borisy, M. Emerman, and T. J. Hope. 2002.** Visualization of the intracellular behavior of HIV in living cells. *J. Cell Biol.* 159:441-452.
66. **Miller, V. and B. A. Larder. 2001.** Mutational patterns in the HIV genome and cross-resistance following nucleoside and nucleotide analogue drug exposure. *Antivir. Ther.* 6 Suppl 3:25-44.
67. **Mousnier, A., N. Kubat, A. Massias-Simon, E. Segeral, J. C. Rain, R. Benarous, S. Emiliani, and C. Dargemont. 2007.** von Hippel Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step. *Proc. Natl. Acad. Sci. U. S. A* 104:13615-13620.

68. **Moutouh, L., J. Corbeil, and D. D. Richman. 1996.** Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc. Natl. Acad. Sci. U. S. A* 93:6106-6111.
69. **Nowak, M. 1990.** HIV mutation rate. *Nature* 347:522.
70. **Palma, A. C., F. Araujo, V. Duque, F. Borges, M. T. Paixao, and R. Camacho. 2007.** Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.* 7:391-398.
71. **Parissi, V., C. Calmels, V. R. De Soultrait, A. Caumont, M. Fournier, S. Chaignepain, and S. Litvak. 2001.** Functional interactions of human immunodeficiency virus type 1 integrase with human and yeast HSP60. *J. Virol.* 75:11344-11353.
72. **Pommier, Y., A. A. Johnson, and C. Marchand. 2005b.** Integrase inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.* 4:236-248.
73. **Pommier, Y., A. A. Johnson, and C. Marchand. 2005a.** Integrase inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.* 4:236-248.
74. **Potter, S. J., C. B. Chew, M. Steain, D. E. Dwyer, and N. K. Saksena. 2004.** Obstacles to successful antiretroviral treatment of HIV-1 infection: problems & perspectives. *Indian J. Med. Res.* 119:217-237.
75. **Rhee, S. Y., T. F. Liu, M. Kiuchi, R. Zioni, R. J. Gifford, S. P. Holmes, and R. W. Shafer. 2008.** Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology.* 5:74.
76. **Rice, P., R. Craigie, and D. R. Davies. 1996.** Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* 6:76-83.
77. **Shun, M. C., J. E. Daigle, N. Vandegraaff, and A. Engelman. 2007.** Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular emerlin protein. *J. Virol.* 81:166-172.
78. **Smolov, M., M. Gottikh, V. Tashlitskii, S. Korolev, I. Demidyuk, J. C. Brochon, J. F. Mouscadet, and E. Deprez. 2006.** Kinetic study of the HIV-1 DNA 3'-end processing. *FEBS J.* 273:1137-1151.
79. **Turechek, W. W. 2004.** Nonparametric tests in plant disease epidemiology: characterizing disease associations. *Phytopathology* 94:1018-1021.
80. **Turechek, W. W. and L. V. Madden. 2000.** Analysis of the association between the incidence of two spatially aggregated foliar diseases of strawberry. *Phytopathology* 90:157-170.

81. **Wain-Hobson, S., C. Renoux-Elbe, J. P. Vartanian, and A. Meyerhans. 2003.** Network analysis of human and simian immunodeficiency virus sequence sets reveals massive recombination resulting in shorter pathways. *J. Gen. Virol.* 84:885-895.
82. **Wiskerchen, M. and M. A. Muesing. 1995.** Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J. Virol.* 69:376-386.
83. **Yoder, K. E. and F. D. Bushman. 2000.** Repair of gaps in retroviral DNA integration intermediates. *J. Virol.* 74:11191-11200.
84. **Yusa, K., M. F. Kavlick, P. Kosalaraksa, and H. Mitsuya. 1997.** HIV-1 acquires resistance to two classes of antiviral drugs through homologous recombination. *Antiviral Res.* 36:179-189.
85. **Zhu, K., C. Dobard, and S. A. Chow. 2004.** Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J. Virol.* 78:5045-5055.
86. **Zouhiri, F., J. F. Mouscadet, K. Mekouar, D. Desmaele, D. Savoure, H. Leh, F. Subra, M. Le Bret, C. Auclair, and J. d'Angelo. 2000.** Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J. Med. Chem.* 43:1533-1540.
87. **Vallari, J Yamaguchi, V Holzmayer, B Harris, C Ngansop, D Mbanya, N Ndembi, L Kaptue, S Devare, and Catherine Brennan,** HIV-1 Group P: Confirmation of Group P in Cameroon: <http://retroconference.org/2010/>